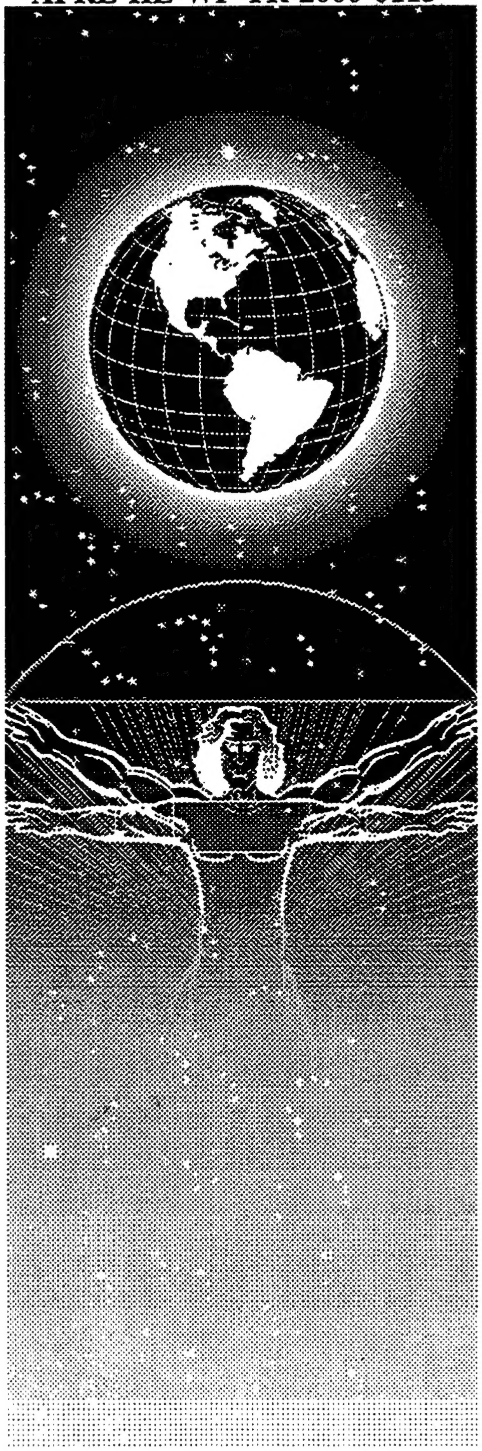
AFRL-HE-WP-TR-2000-0115



UNITED STATES AIR FORCE
RESEARCH LABORATORY

EMPIRICAL INVESTIGATIONS OF TRUST-RELATED SYSTEMS VULNERABILITIES IN AIDED, ADVERSARIAL DECISION MAKING

Ann Bisantz
James Llinas
Younho Seong
Richard Finger
Jiun-Yin Jian

CENTER FOR MULTISOURCE INFORMATION FUSION
DEPARTMENT OF INDUSTRIAL ENGINEERING
STATE UNIVERSITY OF NEW YORK AT BUFFALO
BUFFALO NY 14260

MARCH 2000

INTERIM REPORT FOR THE PERIOD OCTOBER 1998 TO SEPTEMBER 1999

20010501 134

Approved for public release; distribution is unlimited

Human Effectiveness Directorate
Crew System Interface Division
2255 H Street
Wright-Patterson AFB OH 45433-7022

NOTICES

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Please do not request copies of this report from the Air Force Research Laboratory. Additional copies may be purchased from:

National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22161

Federal Government agencies and their contractors registered with the Defense Technical Information Center should direct requests for copies of this report to:

Defense Technical Information Center
8725 John J. Kingman Road, Suite 0944
Ft. Belvoir, Virginia 22060-6218

TECHNICAL REVIEW AND APPROVAL

AFRL-HE-WP-TR-2000-0115

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

MARIS M. VIKMANIS
Chief, Crew System Interface Division
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
March 2000

3. REPORT TYPE AND DATES COVERED
Interim Report: October 1998 - September 1999

4. TITLE AND SUBTITLE
Empirical Investigations of Trust-Related Systems Vulnerabilities in Aided, Adversarial Decision Making

5. FUNDING NUMBERS
C: F41624-94-D-6000
PE: 62202F
PR: 7184
TA: 10
WU: 46

6. AUTHOR(S)
Ann Bisantz, James Llinas, Younho Seong, Richard Finger, Jiun-Yin Jian

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Center for Multisource Information Fusion
Department of Industrial Engineering
State University of New York at Buffalo
Buffalo, NY 14260

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory
Human Effectiveness Directorate
Crew System Interface Division
Air Force Materiel Command
Wright-Patterson AFB OH 45433-7022

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-HE-WP-TR-2000-0115

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

In modern military environments, command-and-control decisions are increasingly supported by information systems which collect, analyze, and display information from multiple sources and sensors, to give decision-makers real time information about an evolving tactical situation. Aided-adversarial decision-making (AADM) refers to military command and control decision in such environments, in which computerized aids are available to groups of co-located and distributed decision-makers, and in which there is a potential for adversarial forces to tamper with and disrupt such aids. In aided, adversarial, decision-making environments, various threats from and offensive opportunities for Information Warfare (IW) activities may exist. In these situations, it is crucial to understand the effect of degraded or altered information on human decision-makers, particularly when that information may be intentionally manipulated. The research described in this report continues two prior phases of research which focused on defining, characterizing, and (where possible) modeling the dependencies and vulnerabilities of AADM on components of information, and considered the role of the human decision maker in AADM, developing a theoretical framework to investigate issues of trust in AADM, and a scale to measure human-automation trust. This report presents research which describes and further develops the theoretical approach begun earlier, describes the completed trust scale, and describes an experimental test bed and an initial experiment which tested the theoretical framework developed. Additionally, an initial description of how cultural issues in AADM can be represented by formalisms in different decision-making models is presented, and experimentation in the area of graphical data presentation and trust in AADM is described.

14. SUBJECT TERMS
information warfare, trust in decision aids, uncertainty visualization, adversarial decision-making

15. NUMBER OF PAGES
95

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFED	UNCLASSIFIED	UNL

This page left blank intentionally.

PREFACE

This effort was accomplished under Contract F41624-94-D-6000, Delivery Order 0007 for the Air Force Research Laboratory's Human Effectiveness Directorate, Crew System Interface Division, Information Analysis and Exploitation Branch (AFRL/HECA). It was completed for the prime contractor, Logicon Technical Services, Inc. (LTSI), Dayton Ohio, under Work Unit No. 71841046: "Crew Systems for Information Warfare." Mr. Don Monk was the Contract Monitor and Mr. Gilbert Kuperman was the Technical Monitor.

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.0 Problem Statement

In modern military environments, command-and-control decisions are increasingly supported by information systems which collect, analyze, and display information from multiple sources and sensors, in order to give decision-makers real time information about an evolving tactical situation. Aided-adversarial decision-making (AADM) refers to military command and control decision in such environments, in which computerized aids are available to groups of co-located and distributed decision-makers, and in which there is a potential for adversarial forces to tamper with and disrupt such aids. In aided, adversarial, decision-making environments, various threats from and offensive opportunities for Information Warfare (IW) activities may exist. In these situations, it is crucial to understand the effect of degraded or altered information on human decision-makers, particularly when that information may be intentionally manipulated.

To date, we have completed two phases of investigation into these issues. We consider AADM to be representative of the modern-day military environment, in which each decision-maker is supported by some type of automated, decision-aiding software system. The effectiveness of the synergy between the human operator and computer systems is clearly influenced by many factors, such as the respective sophistication of each adversary's decision-aiding technology, the underlying cultural differences between the adversaries, and potential mismatches between the automation design and operator expectations.

1.1 Prior Research – Phase 1

The nature of several of these factors were examined in our phase one effort (Llinas, Drury, Bialas, and Chen, 1997). In this phase, we focused on the original tasks of defining, characterizing, and (where possible) modeling the dependencies and vulnerabilities of AADM on components of information. In the research we are conducting, the decision-aiding technology is assumed to be based on data fusion processes. The central theme of this work involved characterizing the value of information to the decision-maker, and emphasized a variety of quantitative models and descriptions of the normative value of information in decision-making. This study also examined cultural effects, human error patterns, and other factors influencing the potential value of information. While the sociological and human factors engineering literature address some of these issues in various contexts, there has been little work examining the relative importance of these factors on overall performance in an AADM environment. As comprehensive empirical investigations into the interactions and effects of all of these factors were beyond the scope of our ongoing studies, we hypothesized that an examination of human trust in the automated aid could provide critically useful information in a more straightforward manner. This formed the basis for the second phase of our investigation.

1.2 Prior Research – Phase 2

In the second phase, we focused on the role of the human decision maker in AADM. In particular, we investigated issues of human trust in the decision aid, in cases where the aid would be based on data fusion processes. We postulated that trust is important in AADM environments, from both an offensive and defensive perspective, and that understanding trust-based influences on decision-making performance could be a critical issue in such situations. From an offensive perspective, IW techniques could be used to distort the information provided to an enemy by their own information and decision aiding systems, disrupting their trust in and probably use of such systems. Alternatively, one might want to intentionally deceive an enemy, and would want him to continue trusting the altered information. Similarly, friendly forces might be vulnerable to information attacks which diminish their trust in data fusion or other decision aids, rendering these assets less useful, or to deceptive attacks, in which an inappropriately high level of trust in the aid is maintained.

In both of these cases, it is important to understand the components of trust, how trust changes over time, how trust affects action, and how people might "calibrate" their trust to an appropriate level. In the second phase of our research, we investigated research on trust from sociological and engineering perspectives, the former focusing on trust between people (e.g., Rempel, Holmes, and Zanna, 1985), and the latter on trust between humans and automation or systems (e.g., Muir and Moray, 1996). From these investigations, we concluded that trust is best seen as a multi-dimensional construct, reflecting a set of interrelated perceptions (e.g., the reliability, or predictability of an entity) and actions (e.g., use of an automated system, reliance on a person). The concept of trust is based on past experience of a person with the entity to be trusted, characteristics of the entity (e.g., is it predictable, are its mechanisms understandable), and characteristics of the person (e.g., in the case of automated systems, someone might "overtrust" the automation if he lacks the skills to take over manually). Trust also has dynamic characteristics (Lerch and Prietula, 1989), changing over time, and in response to specific events. It can increase over time as people have experience with a reliable and predictable system, and then degrade if that system exhibits faulty behavior. Once trust has degraded, it can take time before trust in the system can be regained. Trust as a construct is important in AADM environments insofar as it impacts changes over time in the behavior of a decision-maker, with respect to their use of information and decision aids.

During the second phase of research, in keeping with the multi-dimensional nature of trust, we began the development of an empirically based, multi-dimensional scale of trust, which could be used to measure individual's feelings of trust for automated systems such as decision aids. We also developed a framework for organizing experimentation in this area, based on the properties of the AADM environment, and characteristics of potential information warfare attacks. Specifically, we identified four potential dimensions which could be varied in experimentation: system level, surface-depth level, malfunction type, and cause of corruption. System level refers to the point at which the information attack could occur: from the sensors in the environment, to the decision aiding algorithms, to the display. The surface-depth level refers to the point within the system levels where the corruption takes place: either within the system mechanism (e.g., the data fusion algorithm) or its output. The level of malfunction encompasses point failures to constant and random errors, and the cause of corruption ranges from unintentional to sabotage and subterfuge.

As we identified in the second phase of our research, there has been some limited empirical work done in the area of human trust in automated systems (e.g., Lee and Moray, 1994; Muir and Moray, 1996; Singh, Molloy, and Parasuraman, 1993). However, this limited work has not addressed variations across all the dimensions of our proposed experimental framework described above. Most importantly, there has not been prior experimental work in the area of human trust which addressed *adversarial* degradation or manipulation of system components and information sources. We believe that people's reliance on information and automated algorithms, and in particular, the event- and time-driven patterns of trust development and diminishment may vary significantly between adversarial and non-adversarial situations.

1.3 Phase 3

1.3.1 Goals

The goals of the current research phase were oriented towards the initial empirical investigation of trust-related vulnerabilities in AADM. The second phase work began to set the stage for these empirical investigations, by defining both metrics and appropriate experimental methodologies for the conduct of unique, exploratory experiments which characterize the nature and effects of trust in AADM situations.

The goals of the current research included the development of appropriate experimental environments, the completion and implementation of a trust metric begun in Phase 2, and the performance of initial pilot experiments. An additional goal included further investigation into cultural aspects of AADM, to continue work begun in Phase 1 of this research.

1.3.2 Report Overview

The remainder of this report is organized as follows: Chapter 2 describes and further develops the theoretical approach begun in Phase 2, as well as describes the completed trust scale. Chapter 3 presents an initial description of how cultural issues in AADM can be represented by formalisms in different decision-making models. Chapter 4 described the experimental test bed that has been developed, and Chapter 5 describes the pilot experiment that was conducted. Chapter 6 describes work done in the area of graphical data presentation and trust in AADM. Finally, Chapter 7 presents goals for future work.

CHAPTER 2

STUDYING PERFORMANCE IN AIDED ADVERSARIAL DECISION MAKING

2.1 Introduction

Aided-adversarial decision-making (AADM) refers to military command and control decision in environments in which command-and-control decisions are increasingly supported by information systems which collect, analyze, and display information from multiple sources and sensors, in order to give decision-makers real time information about an evolving tactical situation. Additionally, the adversarial nature of the environments creates the situation in which there is a potential for adversarial forces to tamper with and disrupt such aids.

Decision aids based on information or data fusion technologies are one potential tool for supporting commanders in tactical situations. Data fusion can be described as a multi-level process by which data from multiple sensors are combined, through a variety of mathematical techniques, to obtain meaningful information not available from one source alone (Walts and Llinas, 1990). In a military context, data fusion has been identified as a means to perform assessments of identities, situations, and threat potential based on information derived from multiple electronic and intelligence sources. In these situations, the inherent risks, time pressure and large volume of data have led to the need for computerized aids performing automated data fusion (Walts and Llinas, 1990). The process of data fusion includes multiple levels, each of which represents a different level of abstraction of the data. These levels include low-level processing (Level 1), in which potential targets and their characteristics are detected and identified, to higher level (Level 2) association of targets into organized groups with certain behaviors, to a third level estimating the potential threat of those groups. Thus, the results of data fusion processing can provide input to the situation assessment activities of battlefield commanders. Additionally, there is a fourth level of processing, by which the data fusion process itself can be controlled or adapted by selection and optimization of multiple potential algorithms for a particular data fusion process (Llinas, Drury, Bialas, and Chen, 1998). Ultimately, information resulting from the data fusion process is presented to the human decision-maker through a computer interface. This process model, originally developed by the Joint Directors of Laboratories Data Fusion Group (JDL/DFG), a defense laboratory data fusion technology oversight committee, is illustrated in Figure 1.

Due to the adversarial nature of the environment, hostile forces may attempt to compromise tactical decision-making through "Information Operations" or "Information Warfare." Information operations include a wide range of activities intended to affect information systems and subsequent decisions, from large scale deceptions to targeted information attacks. Information attacks include actions taken to disrupt or alter an adversary's information systems without physically compromising the system, making detection of such manipulations difficult (Llinas et al., 1998). Examples of these attacks

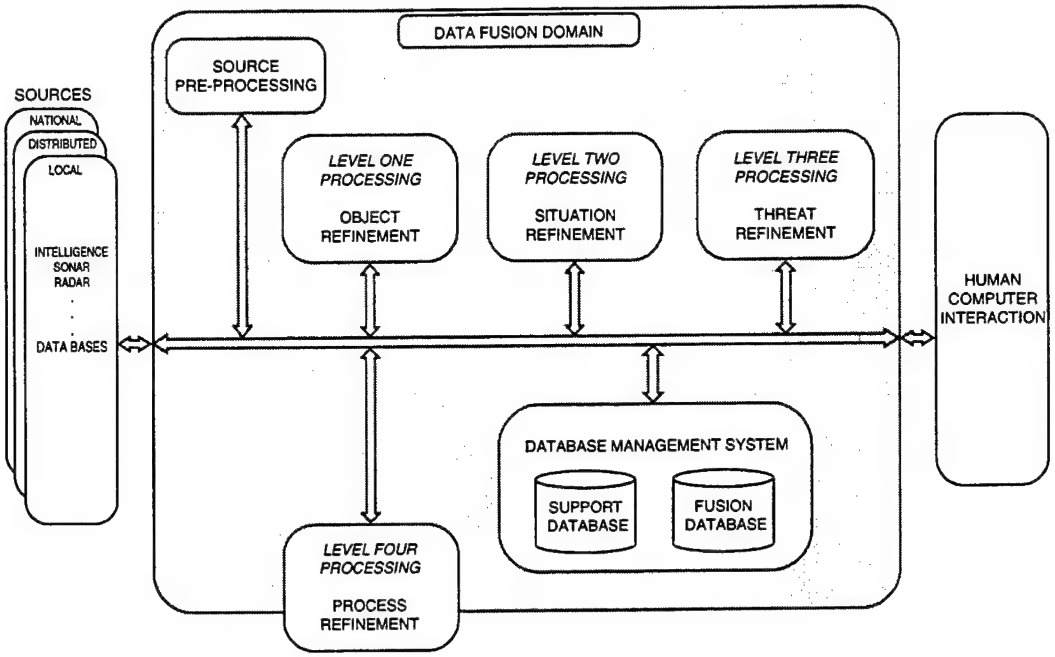


Figure 2.1. Model of Data Fusion Processing (Llinas, Drury, Bialas, and Chen, 1998).

include such things as the interjection of erroneous information into electronic data streams, or computer virus attacks on information systems.

In AADM, the primary concern is defense against information attacks promulgated by adversarial forces, rather than taking offensive actions against an adversary. Data fusion based decision aids, due to their reliance on the integrity of both sensed and transmitted information, and the information-processing algorithms that produce the "fused" information or situation estimates, are potential targets for information warfare. Adversaries, through information attack, could corrupt data sources, data as it is being transmitted from the source to the information system, or between the information systems of multiple decision-makers, or the data fusion system itself.

As further developed in Phase 2 of this research, given the potential for information operations to disrupt and corrupt information provided by data-fusion based aids, it is necessary to understand the extent to which decision-makers rely on or use these aids, and factors affecting that reliance. A possible source of information regarding these issues is research that has been performed in the area of human trust in automated systems (e.g., Lee and Moray, 1992; Muir and Moray, 1996; Parasuraman, Molloy, and Singh, 1993; Sheridan, 1988). Researchers have suggested that trust can affect how much people accept and rely on increasingly automated systems (Sheridan, 1988).

Generally, research from both social science and engineering perspectives agree that trust is a multi-dimensional, dynamic concept capturing many different notions. For example, Rempel et al. (1985) concluded that trust would progress in three stages over time from predictability, to dependability to faith. Muir and Moray (1996) extended these three factors, and developed an additive trust model that contained six components: predictability, dependability, faith, competence, responsibility, and reliability. Sheridan (1988) also suggested possible factors in trust, including reliability, robustness, familiarity, understandability, explication of intention, usefulness, and dependence.

Empirical results have shown that people's strategies with respect to the utilization of an automated system may be affected by their trust in that system. For example, Muir and Moray (1996) and Lee and Moray (1994) studied issues of human trust in simulated, semi-automated pasteurization plants. These studies showed, among other results, that operators' decisions to utilize either automated or manual control depended on their trust in the automation and their self confidence in their own abilities to control the system. Additionally, results showed that trust depended on current and prior levels of system performance, the presence of faults, and prior levels of trust. For example, trust declined, but then began to recover, after faults were introduced (Lee and Moray, 1992). Lerch and Prietula (1989) found a similar pattern in participants' confidence in a system for giving financial management advice: confidence declined after poor advice was given, then recovered, but not to the initial level of confidence.

In the context of AADM, there exists the potential for several circumstances in which trust in data-fusion based decision aids could be affected. For instance, information warfare techniques could be used by an adversary to distort the information provided by decision aiding systems, disrupting (appropriately) commanders' trust in, and utilization of, such systems. Alternatively, an adversary might act deceptively, fooling a commander into trusting and acting based on information in a way favorable to the adversary. Finally, an adversary might disrupt a commander's trust in an aid that is providing good ("trustworthy") information. Understanding the dynamic characteristics of trust development, and loss of trust in response to system events, as studied by Lerch and Prietula (1989), is therefore important in AADM environments. In summary, decision-makers might be vulnerable to information attacks which diminish their trust in data fusion or other decision aids, rendering these assets less useful, or to deceptive attacks, in which an inappropriately high or low level of trust in the aid is maintained. For these reasons, it is necessary to investigate human trust in AADM situations, in order to better understand how data-fusion based decision aids will impact the decision-making process under different circumstances.

2.2 Theoretical Framework

To investigate human trust in AADM, Phase 2 began the development of a theoretical framework. This framework was further developed in the present phase. Understanding the impact of such corruption on resultant decisions requires an analysis of potential factors which may influence characteristics of the corruption. To this end, a framework was developed to integrate and systematically vary the factors which could influence human performance in AADM environments. These factors are drawn in part from an examination of some of the experimental studies on human trust in automated systems cited above.

1. Locus of Attack. One potential factor is the location at which the potential for corruption exists. Two potential dimensions can contribute to this factor: the component dimension, and the surface-depth dimension.

a) Component Dimension. Information could be corrupted at a variety of components, or levels, in the AADM environment. Information could be corrupted at the level of the tactical situation (by interfering with sensors), within the information processing and data fusion algorithms that comprise the decision aids, or at the level of the human-computer interface.

- Environment. The physical environment in AADM corresponds to the actual tactical situation that is taking place. Just as the states of pumps and heaters can be observed and controlled, the states (e.g., current locations, available weapons) of hostile and friendly assets can be assessed and sensed, and actions related to the situation can be taken.
- Automated Aiding System. The level comprising the data fusion algorithms and processes, which automatically combine and synthesize information obtained from the tactical environment to form the basis for control or decision actions, can be considered analogous to the automated controller which used information from the physical control system to take control actions. However, in AADM, the data fusion process itself consists of multiple levels, as illustrated in Figure 1. Information could be corrupted at any or all of the levels of data fusion processing.
- Interface. Finally, in an AADM environment, one can consider a third, interface level in which the results of the data fusion algorithms are displayed to the operator, in order to aid decision making. Inaccurate results, or manipulated displays, could be presented to the operator.

b) Surface-Depth Dimension. A second related dimension along which investigations of performance in AADM systems can vary is a surface-depth dimension. The surface level corresponds to the information available about the environment (as formalized in Brunswik's Lens Model; Cooksey, 1996; Hammond, Stewart, Brehmer, and Steinman, 1975), whereas the depth level corresponds to the actual state of the environment. In an AADM environment, surface level features would be the observable outputs from sensors, or data fusion processes. Depth level features would be the actual operations of the sensors or algorithms themselves. This surface-depth dimension can be applied at the environment and aiding system dimensions described above, resulting in five combinations, as described in Table 1. Since the interface level does not perform any processing, but is a representation of the results of the aiding system level, it is best considered at the surface level.

2. Malfunction Level. Information aids for AADM can fail or be corrupted in qualitatively different ways, either failing completely, or being partially degraded, resulting in two malfunction levels:

Table 2.1. Components Of An Aided Adversarial Decision Making Environment Described Along A System And A Surface-Depth Dimension.

	Surface-Depth Dimension	
System Dimension	Surface Level	Depth Level
Environment Level: Tactical Situation	Sensed and Observed Data	Evolving Tactical Situation
Intervention Level: Data Fusion Algorithms	Results of Algorithms	Data Fusion Algorithms
Interface Level: Decision Aid	Display Format	n/a

- Element failure. System components can fail completely resulting in a loss of data.
- Element degradation. The quality of information provided by the system component can be degraded, resulting in partial information loss, or increased ambiguity and uncertainty.

3. Causes of Failure or Corruption. Information can be corrupted through different causes or intentions, ranging from naturally occurring system failures (e.g., hardware malfunctions), to deliberate attacks on the information systems, to deliberate attacks which are disguised by the adversary.

4. Time Patterns of Failure. A final dimension reflects the dynamic or time-dependent characteristics of the degradation. Failures, sabotage, and subterfuge can occur not only as failures or degradations at a particular point in time, but also in a continuing fashion. Additionally, failures can occur with patterns that are either predictable or unpredictable.

2.3 Framework for Human Factors Research in AADM

The combination of these factors: locus of corruption (component and surface depth levels), malfunction level, causes of corruption, and time patterns of failure, provides a framework, illustrated in Figure 3. As discussed in the next section, this framework can be used to organize future investigations in the area of decision making under aided, adversarial conditions.

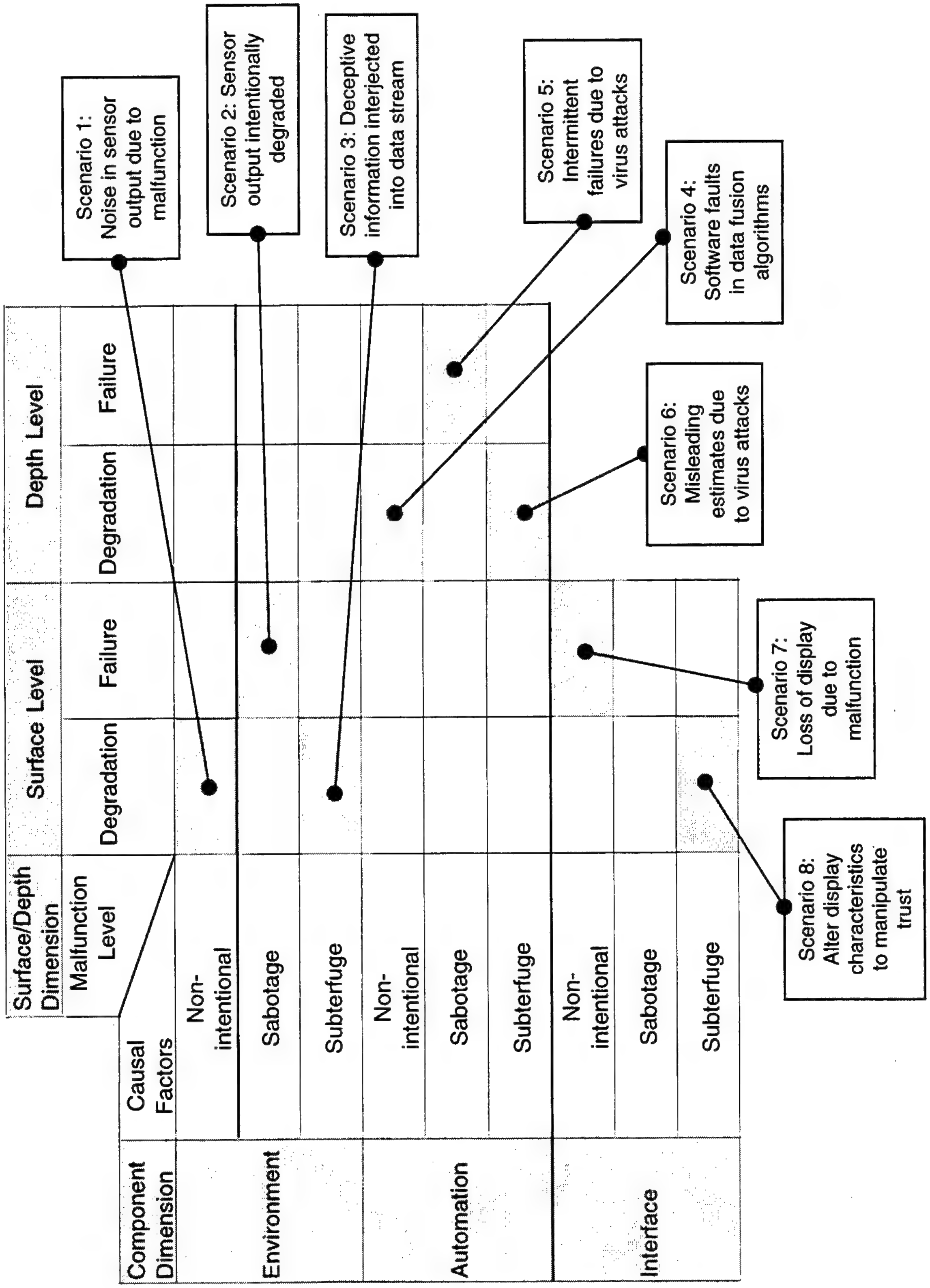

Component Dimension
Causal Factors
Surface/Depth Dimension
Malfunction Level
Surface Level
Depth Level
Degradation
Failure
Degradation
Failure
Environment
Automation
Interface
Non-intentional
Sabotage
Subterfuge
Non-intentional
Sabotage
Subterfuge
Non-intentional
Sabotage
Subterfuge
Scenario 1: Noise in sensor output due to malfunction
Scenario 2: Sensor output intentionally degraded
Scenario 3: Deceptive information interjected into data stream
Scenario 5: Intermittent failures due to virus attacks
Scenario 4: Software faults in data fusion algorithms
Scenario 6: Misleading estimates due to virus attacks
Scenario 7: Loss of display due to malfunction
Scenario 8: Alter display characteristics to manipulate trust

Figure 2.2. Framework for experimentation for multi-crew AADM.

2.4 Potential Scenarios

Given a particular experimental context, the framework described above can be used to systematically define a series of experimental manipulations in order to investigate individual and group decision-making performance.

For example, for an evolving tactical situation, the depth dimension corresponds to the actual states and activities of the various player. In turn, the degree or nature of this tactical environment-depth factor could be varied over levels such as "Benign" or "Threatening" or "Critical." The surface level in this case corresponds to sensed or observed information about the environment (i.e., the tactical situation). Again, this tactical-surface factor could be varied over levels such as "Uncorrupted" or "Moderately Corrupted" or "Severely Corrupted."

At the automation level (the data fusion algorithms), the depth level of the surface-depth dimension corresponds to the structure of actual data fusion algorithms and procedures themselves, comprising all four levels of the data fusion process. The surface level reflects the estimates produced by these algorithms.

Finally, at the interface level, the depth dimension corresponds to the actual information or advice that is to be given to the operator (the "state" of the display), while the surface dimension corresponds to the manner or format in which it is displayed. Within these combinations, illustrated in Table 1, levels of the additional factors can be manipulated to form the framework shown in Figure 3. This framework can be used to structure experimental studies or assist in the interpretation of empirical results. Although the final manifestation of failures or disruptions in these different categories may be similar (i.e., degraded estimates displayed by the decision aid), the methods available to detect and circumvent the disruptions differ. For example, consider the disruption of sensors in the environment to interject deceptive information into data being transmitted by the sensors. This disruption may mimic, at the decision-aid level, actual tactical deceptions by adversarial forces. Checks of other data sources (e.g., other sensors, visual reports) might reveal discrepancies pointing to the former deception, but remain consistent with the latter types of deception. Similarly, faults in the data fusion algorithms themselves would remain constant across different sources of data, in contrast to situations when only particular inputs to the data fusion processes were corrupted.

2.5 Dependent Measures

To study human performance in the above scenarios, there are several categories of dependent measures which can be used. Performance measures, such as judgment accuracy and time to act, can be used to assess participants' overall success in executing the scenarios. Process measures, such as the types of information sources accessed, or the timing and sequence of decisions or actions can be used to assess differences in strategy that may result from different experimental conditions. Finally, subjective trust measurement scales can be used to assess aspects of participants' trust in the AADM system.

2.5.1 Trust Measurement Scale

Phase 2 of this research included the beginning development of a measurement scale to assess human trust in computerized systems. This development has been completed (Jian, 1998; Jian, Bisantz, and Drury, 2000), and a brief summary of the methodology used to develop the scale, and the resultant scale which was utilized in this research phase, will be presented here. For further detail see Jian (1998).

2.5.1.1 Background.

Past research on trust from both engineering and sociological perspectives have utilized questionnaires or measurement scales to assess trust. Although the questionnaires are similar in that they have treated trust as a multi-dimensional concept, the factors of trust, and thus the attributes and descriptors included in the questionnaires, have been based on different theoretical notions of trust, depending on the theoretical orientation of the researcher. For example, Rempel et al. (1985) concluded that trust would progress in three stages over time from predictability, to dependability to faith. Muir and Moray (1996) extended these three factors, and developed an additive trust model that contained six components: predictability, dependability, faith, competence, responsibility, and reliability. Sheridan (1988) also suggested possible factors in trust, including reliability, robustness, familiarity, understandability, explication of intention, usefulness, and dependence. Additionally, the questionnaires used in past research differ in that some are designed to measure trust in a particular person or system, while others measure a more general, non-directed propensity to be trusting. For example, Larzelere and Huston (1980) and Rempel et al. (1985) designed questionnaire items that measured trust in a specific individual (a romantic partner), and Lee and Moray (1996) asked questions specific to the control of an experimental system. In contrast, work by Singh et al. (1993) addressed a general potential for complacency by using questionnaire items about a variety of automated systems. Finally, previous studies have generally assumed that the concepts of trust and distrust were opposites. It could be that these concepts (trust and distrust) in fact encompass very different types of concepts or factors, as for example, do the concepts of comfort and discomfort (Zhang, Helander, and Drury, 1996).

Given the current state of research on trust measurement, several assertions can be made. First, as noted above, the questionnaires used to measure trust have included items based on different theoretical notions of trust, and have not been based on an empirical analysis which attempted to uncover multiple components of trust. Second, the previous studies have not explicitly evaluated how trust between human and automated systems differs from trust between humans, or for that matter, from trust in general. Although researchers in human-machine systems have employed concepts of trust from sociological studies, there is no empirical basis for necessarily assuming that concepts of human-machine trust are identical to trust between humans. Were such differentiated scales developed, they could provide a potentially more reliable and valid tool for assessing people's trust in automated, computerized systems.

2.5.1.1 Method.

To address these issues, a three-phased experimental study was conducted of the concept of trust by an individual in another individual or system. The goal of these experiments was to explore the underlying factors comprising the concepts of trust, and to develop a potentially more reliable and valid tool for assessing people's trust in automated systems. The experiments are modeled after those conducted by Zhang, Helander, and Drury (1996) who developed a measurement scale for the similarly complex notion of comfort.

In the first phase, a word elicitation study, we collected various words related to concepts of trust and distrust. In the second phase, a questionnaire study, we investigated how closely each of these words was related to trust or distrust in order to evaluate whether or not trust and distrust were opposites or represented somewhat different concepts, and whether or not concepts of trust and distrust were similar for general trust, trust between people, and trust between humans and systems. The third phase was a paired comparison study, in which participants rated the similarity of pairs of words. Data from both the questionnaire study and the paired comparison study were then used to construct a multi-dimensional measurement scale for trust.

2.5.1.2 Word Elicitation Study.

The objective of this phase was to collect a large set of words related to trust and distrust. Seven graduate students majoring in Linguistics or English were asked to provide written descriptions of their understanding of both trust and distrust with respect to either trust between people, trust in automation, or trust with no specific qualification. Next, participants were also asked to rate whether a set of 138 words were related to trust using a nominal scale, with "positively related to trust," "not related to trust," "negatively related to trust," and "don't know" as scale points. This initial set of 138 words was collected by analyzing questionnaires used in previous studies, and from dictionary definitions and thesauri. As with the written descriptions, these ratings were performed with respect to the three conditions of trust between people, trust in automation, and general trust.

Thirty-eight new words from the written descriptions of trust provided by participants' questionnaires. In addition, we eliminated words from the initial set based on participants' ratings of the words as either not-related to trust, or having ambiguous ratings. To provide continuity with the existing literature, words retrieved from questionnaires used in previous research were not eliminated, although some were rated as "not related to trust" (e.g., familiarity). A total of 64 words were eliminated. After eliminating these words and adding the new words, the final set of words contained 112 trust-related words, which were used in the subsequent questionnaire study.

2.5.1.3 Questionnaire Study.

In this experiment, 125 participants were asked to rate the extent to which words from Set-1 were related to trust or distrust, from the perspective of either trust in general, or trust between people, or trust in automated systems, for a total of six between-subject conditions. Participants rated the relatedness of the word to trust or distrust using a seven point scale, with end points of "positively related to trust (or distrust)" and "negatively related to trust (or distrust)."

Based on participants' ratings, several conclusions were made. First, based on average ratings of trust or distrust, correlations between the average ratings of trust and distrust, for all three conditions, were highly negatively correlated ($r = -.96$, $r = -.95$, $r = -.95$, for general trust, human-human trust, and human-machine trust), indicating that concepts of trust and distrust could be treated as opposites. Additionally, we compared ratings of individual words across the three conditions of general, human-human, and human-machine trust, to see how individual words might be differently related to the three types of trust.

Words were assigned, according to their average ratings, into the top 5, 10, 15, 20, 25, and 30 words most related to trust and distrust, for each condition. For example, the five words most related to general trust were *trustworthy*, *honesty*, *loyalty*, *reliability*, and *honor*. The five words most related to trust between humans and automated systems were *trustworthy*, *loyalty*, *reliability*, *honor*, and *familiarity*. The five words most related to trust between people were *trustworthy*, *honesty*, *loyalty*, *reliability*, and *integrity*. The degree to which these sets overlap gives an indication of the extent to which concepts of trust and distrust were similar for the three conditions. A comparison of the sets, and the degree of overlap, indicated that the three concepts of trust were similar. Based on the ratings, a set of 15 words most related to trust, and 15 words least related to trust, were selected for investigation in the next phase. These words are shown in Table 2.2.

Table 2.2. Results From the Questionnaire Study

15 Words Least Related to Trust	15 Words Most Related to Trust
Betray	Assurance
Beware	Confidence
Cheat	Entrust
Cruel	Familiarity
Deception	Fidelity
Distrust	Friendship
Falsity	Honesty
Harm	Honor
Lie	Integrity
Misleading	Love
Mistrust	Loyalty
Phony	Promise
Sneaky	Reliability
Steal	Security
Suspicion	Trustworthy

2.5.1.4 Paired Comparison Study.

Thirty participants were asked to rate the similarity between pairs of words found in Table 2.2. All possible pairs were presented to participants (a total of 435 pairwise comparisons). Participants used a computerized rating program to rate each pair of words on a seven-point scale with end points of "Totally different" and "Almost the same" (Zhang et al., 1996) by clicking on the appropriate rating. Cluster analysis was used to group words according to their similarity to each other, as measured in the paired-comparison study. The between-group average linkage method was performed using SPSS. The cluster analysis provided a method to identify clusters of similar words. Twelve clusters were identified (See Table 2.3), and used to create the trust questionnaire.

Table 2.3. Trust Scale Items, for Human-Machine Trust, and the Corresponding Cluster of Trust Related Words on Which They Were Based

Item	Words Groups from Cluster Analysis
The system is deceptive	Deception Lie Falsity Betray Misleading Phony Cheat
The system behaves in an underhanded manner	Sneaky Steal
I am suspicious of the system's intent, action, or output	Mistrust Suspicion Distrust
I am wary of the system	Beware
The system's action will have a harmful or injurious outcome	Cruel Harm
I am confident in the system	Assurance Confidence
The system provides security	Security
The system has integrity	Honor Integrity
The system is dependable	Fidelity Loyalty
The system is reliable	Honesty Promise Reliability Trustworthy Friendship Love
I can trust the system	Entrust
I am familiar with the system	Familiarity

2.5.1.5 Resultant Questionnaire.

Based on the results of the cluster analysis, we developed a proposed trust scale for human-machine trust, which included 12 items for measuring trust between people and automated systems. The 12 items were derived by examining the words in the empirically derived clusters for human-machine trust. Table 2.3 shows the 12 items with respect to groupings of words, while Figure 2.2 shows how the proposed scale might be presented to participants.

The system is deceptive
1 2 3 4 5 6 7

The system behaves in an underhanded manner
1 2 3 4 5 6 7

I am suspicious of the system's intent, action, or outputs
1 2 3 4 5 6 7

I am wary of the system
1 2 3 4 5 6 7

The system's actions will have a harmful or injurious outcome
1 2 3 4 5 6 7

I am confident in the system
1 2 3 4 5 6 7

The system provides security
1 2 3 4 5 6 7

The system has integrity
1 2 3 4 5 6 7

The system is dependable
1 2 3 4 5 6 7

The system is reliable
1 2 3 4 5 6 7

I can trust the system
1 2 3 4 5 6 7

I am familiar with the system
1 2 3 4 5 6 7

Figure 2.3. Proposed questionnaire to measure trust between people and automated systems.

2.6 Summary

In summary, a three-phased experimental study of trust concepts was performed to develop an empirically based scale to measure trust in automated systems. The experiments explored similarities and differences in the concepts of trust and distrust, and among general trust, human-human trust, and human-machine trust. Results provided empirical evidence for considering trust and distrust to be opposites, suggesting that two scales do not need to be developed to measure trust and distrust separately. Additionally, concepts of general trust, human-human trust, and human-machine trust tended to be similar, although people seemed to consider human-human trust more in terms of trust than distrust. Finally, results from the cluster analysis were used to construct a proposed scale to measure trust in human-machine systems.

CHAPTER 3

REPRESENTATION OF AADM CULTURAL ISSUES IN MODELS OF DECISION-MAKING

3.1 Introduction

Cultural issues have been suggested as an important factor in understanding, and subsequently supporting, activities in AADM environments. In particular, an understanding of adversarial patterns of IO interference, or response to friendly IO operations, depends on an understanding of cultural implications. Phase 2 of this effort identified the following definitions of culture, which are related to psychological notions of learned sets of experiences and derivative expectations, shared by groups of individuals (at multiple levels of group size), as well as sociological notions of shared meaning. Culture can exist among a small office group, a business, neighborhood, town, region, country, and across time and political boundaries. From Hoecklin (1995) culture can be described as:

(1) A shared system of meanings. Culture dictates what groups of people pay attention to. It guides how the world is perceived, how the self is experienced and how life itself is organized. Individuals of a group share patterns that enable them to see the same things in the same way and this holds them together. Each person carries within them learned ways of finding meaning in their experiences. In order for effective, stable and meaningful interaction to occur, people must have a shared system of meaning. There must be some common ways of understanding events and behavior, and ways of anticipating how other people in your social group are likely to behave. It is only when the meanings do coincide that effective communication can happen.

(2) It is Relative. There is no cultural absolute. People in different cultures perceive the world differently and have different ways of doing things, and there is no set standard for considering one group as intrinsically superior or inferior to any other. Each national culture is relative to other cultures' ways of perceiving the world and doing things.

(3) It is Learned. Culture is derived from your social environment, not from your genetic make-up.

(4) It is About Groups. Culture is a collective phenomenon that is about shared values and meanings.

One approach to the inclusion of cultural factors in the study of AADM is to consider how cultural issues could be represented in a variety of judgement and decision-making models. Through this assessment, the potential types of impacts that cultural issues could have on decisions, and subsequent actions, could potentially be identified.

For this discussion, the following decision models will be considered: Expected utility theory models, signal detection models, information processing-type models (e.g., Decision-ladder), Lens Models, and recognition-primed decision models.

3.2 Models

3.2.1 Expected Utility Theory Models

Expected utility theory models are normative descriptions of human decision making activity, based on mathematically defensible principles. For example, Simon (1960) describes decision making as a three step process of identifying the need for a decision, identifying alternative courses of action, and then choosing among these alternatives. In this model, it is assumed that actions along with their predicted outcomes can be enumerated and the probabilities of the various outcomes can be determined (Savage, 1954; Lindley, 1985). Action alternatives can then be compared based on their expected benefits, or utilities. These expected utilities are computed by combining the utilities of the outcomes associated with each alternative, and the probability of each outcome. Thus, this model also assumes that the utilities of the various outcomes are consistent, numerically measurable quantities (Von Neumann and Morgenstern, 1953). This model of decision making has been considered to be descriptive of rational decision-making behavior, and therefore a prescription for normatively correct decision making (Savage, 1954; Lindley, 1985).

These models, while not generally considered descriptive of decision-making under conditions of rapidly changing circumstances and time pressure, may be useful in modeling cultural influences on more deliberative decisions, in at least two ways. Consider a circumstance in which A (a friendly commander) is trying to compare which type of information operation to select, to influence an adversary, B. With an expected utility theory model, possible outcomes, and the probabilities of these outcomes, given a decision, are required. Thus, A's understanding of the decision problem would necessarily include an understanding of culturally dependent outcomes or reactions of B, and the probabilities that B would react in each particular way. The potential outcomes and associated probabilities would depend on B's and B's cultural group's past experiences, and expectations on his behavior by other group members.

Alternatively, if A is trying to predict a decision or action to be taken by B, it is necessary for A to understand the utilities assigned by B to different action options. These utilities, often estimated subjectively (i.e., subjective expected utility theory models) can be the result of culturally assigned values or penalties for different options.

3.2.2 Signal Detection Theory Models

Signal detection theory models how a human decides whether a stimuli indicates a "signal," or significant event of interest, or whether the stimuli is due to environmental or background noise. For example, signal detection theory could be used to model whether an operator detects that a display or system has been tampered with, given the pattern of information being displayed. This judgment is molded as a yes/no response, based on the level (amount) of stimuli: if the level is past some threshold point, or criterion value, the stimuli will be judged a signal. Signals and no-signals (noise) are modeled as probability distributions (e.g., normal distributions), and therefore cannot be clearly distinguished by the criterion. Errors (missed signals and false alarms) occur due to the probabilistic nature of the signal and noise distribution. The criterion level is measured by a parameter β, or response bias. Cultural issues affecting a signal detection judgment would be captured through the response bias parameter. A high response bias indicates that stimuli must have a high level before it is judged a signal; the opposite is true for a low β. Thus, the response bias reflects the level or amount of evidence which is necessary to convince a judge that a signal has in fact occurred, which may be affected by cultural biases. A high response bias results in more missed signals (signals whose level does not meet the criterion point), while a low response bias results in more false alarms (noise stimuli whose level is above the criterion level).

3.2.3 Information Processing Models

Information processing models of decision making describe cognitive resources and process steps involved in sensing and interpreting information from the environment, and using that information to make decisions, and take actions. Cognitive resources represented in an information processing approach include working and long term memory, as well as attentional resources. Rasmussen (Rasmussen, Pejetersen, and Goodstein, 1994) describes a decision-making model based on information processing stages called the Decision Ladder. Process stages, and resultant states, are made explicit, and describe a process of activation, observation, state identification, goal selection and option evaluation, task selection, planning, and execution. Short-cuts across stages (e.g., from state identification directly to action execution), based on learned, prior experiences, are also represented in the model.

Cultural influences may be captured in several ways using an information processing approach to decision making. As represented in the decision ladder, sequences of processing stages are heavily influenced by learned experiences. Cultural factors may influence which decisions are deliberated at the goal selection level, which goal options are possible, and how those options are evaluated.

3.2.4 Lens Model

Brunswik's Lens Model with its extensions (Brunswik, 1955; Cooksey, 1996; Hammond, Stewart, Brehmer, and Steinmann, 1975) applies linear models to the description of judgment behavior. As seen in Figure 3.1, the Lens Model is a symmetrical model which describes how both the environmental structure, and patterns of cue utilization collectively contribute to judgment performance. The left side of the figure represents the relationship between cue values and the criterion to be judged ($\mathbf{Y_E}$), while the right side of the figure represents how someone would utilize the cues to make a judgment ($\mathbf{Y_S}$). The judgments and the environmental criterion to be judged are described as linear combinations of environmental cues (with weights $\mathbf{r_{E,i}}$ and $\mathbf{r_{S,i}}$), or available information in the environment. In this way, both the judgment policy and the environmental structure in terms of cue-criterion relationships, are captured. Because the models are based on the same environmental information (the cues), the fit between the model of the human judge and the environmental structure can be formally measured.

Essentially, this allows assessment of the extent to which an individual's judgment policy reflects, or has adapted to, the structure of the environment. Additionally, the extent to which individuals' judgments are consistent with their judgment policy can be quantified. The Lens Model is an ideographic model of decision-making (Cooksey, 1997), representing the judgment policy of an individual decision-maker through the linear judgment model (right side of Figure 3.1), rather than the trends of a group of decision makers. In terms of cultural influences, the mapping to the Lens Model is straight forward. Characteristics of each decision-maker are captured in the weights (**$r_{S,I}$**), indicating the relative contribution of each cue to the judgment, and reflect cultural as well as personal influences on the judgment. Selection of cues, including the exclusion of certain cues, represented by zero weighting on that cue, can also be culturally influenced. Determining cultural influences, as opposed to more personal individual factors, would be possible by comparing models across sets of culturally related decision-makers. Similarities across individual models (e.g., in cue weights, or cue exclusion) could be attributed to culturally or group learned factors, while differences would reflect individual judgment components.

The Lens model also offers a formal way to model cultural understanding. Referring to Figure 3.1, given a set of cues **Xi**, consider the left side of the model as representing the judgment of an adversary given the cue set, and the right side the judgment of a friendly commander trying to mimic the adversary. Given the cues, and the known judgments of both parties, an assessment can be made regarding the degree to which the commander has accurately reflected an adversary's (culturally based and

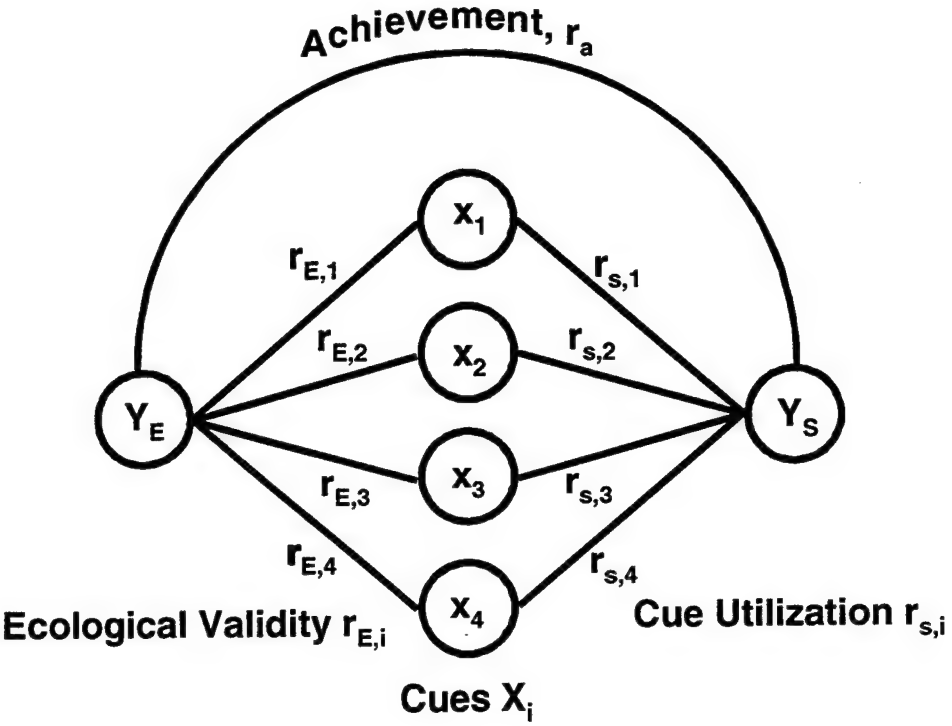


Figure 3.1. Brunswik's Lens Model.

individually based) use of cues. Lens Model parameters to measure this fit can be computed based on correlations between the model predictions for both sides of the model.

3.2.5 Dynamic and Recognition Primed Decision Models

Several researchers (Brehmer, 1990; Brehmer and Allard, 1991; Connolly, 1988; Hogarth, 1981) considered the continuous, cyclical nature of judgment and action in the natural environment under dynamic, time-pressured conditions. In these conditions, immediate feedback about the effects of judgments and actions increases the information available in the environment. Researchers in human-machine systems have provided additional descriptive models of how experienced practitioners actually make decisions in dynamic, time-pressured, real world conditions. For example, in Klein's Recognition-Primed Decision (RPD) model (Klein, 1993; Klein and Calderwood, 1991), decision-makers attempt to recognize and characterize the current situation and identify actions based on their past experience with the situation. They mentally simulate how the actions would work in the current situation, and modify them if necessary. The decision alternatives are compared serially, and the decision-making process ends when an acceptable, not necessarily optimal, solution is reached. The RPD model includės several constructs and mechanisms which could capture cultural influences in AADM. Given a situation that can be mapped to prior experiences, the goals relevant to that situation, the cues that may be important to monitor, expectations about how the situation will evolve, and the set of actions or responses that are appropriate may all be culturally influenced. Assessment of the applicability of past solutions to the current situation, and potential modifications to those solutions, may also be culturally driven.

3.3 Discussion

Describing and modeling cultural issues, due to the need to understand the intentions, potential actions, and reactions of an adversary, are a necessary component to investigating and describing AADM environments. As noted above, culture can be thought of as a shared set of experiences or learned knowledge, across a group of individuals. When applying issues of culture to AADM environments, culture should impact an individuals' decision-making process in a similar manner to other learned information or experiences (i.e., those that are more localized to the individual, rather than group based). Cultural issues are one part of the learned experiences which may influence an individual decision-maker's behavior.

Models of decision-making can capture influences of culture, through the same mechanisms used to capture other experiential influences and resources. Different decision models capture explicitly factors such as subjective value placed on options and probability of reactions (Expected Utility Theory), judgment bias (signal detection theory), goal selection and experience based short cuts in decision processes (decision ladder), influence of environment information on judgements (Lens model) and candidate analogous situations and expectations of situation evolution (RPD). Depending on the particular aspect of AADM being considered (e.g., detecting an IO attack, predicting an enemy response to AADM, or determining a course of action once an IO attack has been detected), different decision models can facilitate the inclusion of relevant cultural influences in the description.

CHAPTER 4

EMPIRICAL APPROACH TO STUDYING TRUST IN AADM

4.1 Application of the Theoretical Framework

Given the experimental framework described in Chapter 2, a variety of baseline experiments could be conducted to investigate issues of human trust in AADM, by selecting and varying levels of different independent factors. For example, the baseline experiments could be conducted to assess the following:

1. The difference in event- and time-driven patterns of trust, again measured through a multi-dimensional rating scale and analysis of relevant actions, when battlefield decision-aiding systems have **different patterns of degradation** (e.g., point failures, consistent failures, consistent degradation, random degradation). For this experiment, the patterns of degradation would be altered. Information could be degraded either by a point failure (e.g., a sensor stops transmitting for a short length of time), a complete failure (e.g., a sensor stops working entirely), consistent degradation (e.g., a sensor will produce a signal with consistent error) and random degradation (e.g., a sensor will produce a signal that contains error that appears random).

2. The difference in event- and time-driven patterns of trust, as measured through a multi-dimensional rating scale and analysis of relevant actions, when battlefield decision-aiding systems are degraded or manipulated **intentionally** (e.g., sabotage), **intentionally with camouflage** (e.g., subterfuge), and **unintentionally**. For this second experiment, which is intended to identify differences in judgment and action behavior based on different causal factors of trust and mistrust, the task instructions given to the participants could be varied. Participants could be told that the probabilistic nature of the information or its unreliability is due to one of three things. In the unintentional condition, participants could be told that information may simply be ambiguous (e.g., both hostile and friendly aircraft could have the same radar profiles), or that some sensors or algorithms may be non-optimal, thus producing information with less than perfect reliability. In the intentional condition, participants could be told that the information may be unreliable due to intentional sabotage on the part of an enemy. In the subterfuge condition, participants could be told that information may be unreliable, and that enemy forces may intentionally sabotage information, and try to conceal their actions. In order to assess only the effects of the sources of degraded information, the same patterns of disrupted information would be present in all three conditions.

3. The difference in event- and time-driven patterns of trust, again measured through a multi-dimensional rating scale and analysis of relevant actions, when battlefield decision-aiding systems are degraded or manipulated at different system levels (e.g., at the level of the sensors, data fusion algorithms, or displays). For the third experiment, uncertainty, or unreliability, would be introduced into the data in one of three levels – at the level of the sensors, the data fusion algorithms, and the displays –

to see how information search strategies, actions, and subjective measures of trust might depend on the locus of uncertainty.

4.2 Validation of Trust Scale

Experiments based on the theoretical framework, such as those described in Section 4.1, present an opportunity to validate the multi-dimensional trust scale which began development during the second phase of our research program. This scale was developed empirically, using methods of paired comparisons and clustering techniques to identify independent factors which contribute to the construct of trust. Thus, the scale should have good internal validity. However, trust as a construct is only valuable if it correlates to action – we are interested in someone's trust in a decision aiding or automation system primarily because trust (or lack of it) may provide an indication of the level or purpose of use of such systems. The experiments described here provide an opportunity to begin to examine the *external or predictive* validity of such a scale; that is, to see how well the measurement of the trust construct maps onto actions or strategies with respect to the fusion-based decision aid. By applying both the multi-dimensional scale, and analysis of participants' actions as dependent measures, it will be possible to assess the success of the scaling measure in predicting participants' actions. If the multi-dimensional scale of trust proved valid in this way, it could prove useful in assessing trust, and thus actions or strategies, in situations where it is more difficult to capture and analyze actions.

4.3 Technical Approach

In order to empirically investigate issues of human trust in AADM situations, it was necessary to construct an experimental test bed. The test bed developed contained the following features:

1. *Battlefield and Data Fusion Simulation.*
   A discrete-event simulation of a battlefield environment, containing multiple, dynamic, and possible uncertain information sources (e.g., position, electronic emissions, weapons capabilities) about potential threats and friendly assets, along with a simulated decision aid based on data-fusion technology was developed using an object-oriented, graphically capable language: Visual **C** ++.

2. *Instrumentation.*
   In order to manipulate the independent variables indicated by the experimental framework developed in phase 2 (briefly described above), the experimental system provides the ability to make several kinds of experimental manipulations. The intent here is to allow experiments in which errors or other types of degradation are introduced into the fusion-based decision aid, in order to measure participants' response (in terms of both trust ratings, and observable actions).

3. Display and Control

To understand the effects of information manipulation and degradation on performance, it was necessary to have an experimental set-up which allows participants to both perceive the possible manipulations, and change their strategies (e.g., take different actions) based on these perceived abnormalities. Therefore the experimental system must allow participants to see the inputs to and outputs from different stages of the decision aid (i.e., through a display), and to take action based on that information (i.e., have some form of control). More specifically, the system allows participants to obtain information from different sources, so they can perceive abnormalities in any one source or stage. Additionally, the system must allow multiple paths of action (e.g., different information search strategies) to assess how decision making strategies might change based on changes in trust in a particular system component.

4. Data Capture

To assess experimental performance, it will be necessary to automatically log participants' interactions with the experimental system.

5. Experimental Scenario

Finally, to provide a context for experimentation, various experimental scenarios, including dynamic sensor and state variable values, and time dependent simulated adversarial or non-adversarial events (e.g., information manipulation or degradation) will be developed.

4.3 Experimental Environment

The experimental environment developed is illustrated in the following figures. The computer display showed aircraft moving in a simulated radar screen window, and information necessary for the task in other text windows. Participants interacted with the system using a mouse to select aircraft and clicking buttons to issue command to the system. Generally, participants saw a map-based radar display window showing various tracks, graphically displayed as hostile, friendly, or unknown. Tracks can be selected by clicking on the track using a mouse.

By selecting from the view menu, either an Information Window or a Data Fusion Window can be selected (See Figure 4.2). The Information Window shows non-data fusion based information about a selected contact: in this case, speed, heading, range, altitude, and radar information. The Data Fusion window shows a simulated data fusion estimate, in the form of a confidence interval, of the probability that a selected track is friendly. These confidence intervals were calculated based on the pre-defined probability of each non-data fusion based information. That is, the interval represented the Bayesian conditional probability of the selected track being friendly based on the pre-defined probabilities of each information. Then, upper and lower confidence interval bounds were randomly selected from 0 to 5 percent, and added (or subtracted, for the lower bound) to the conditional probability for the confidence interval.

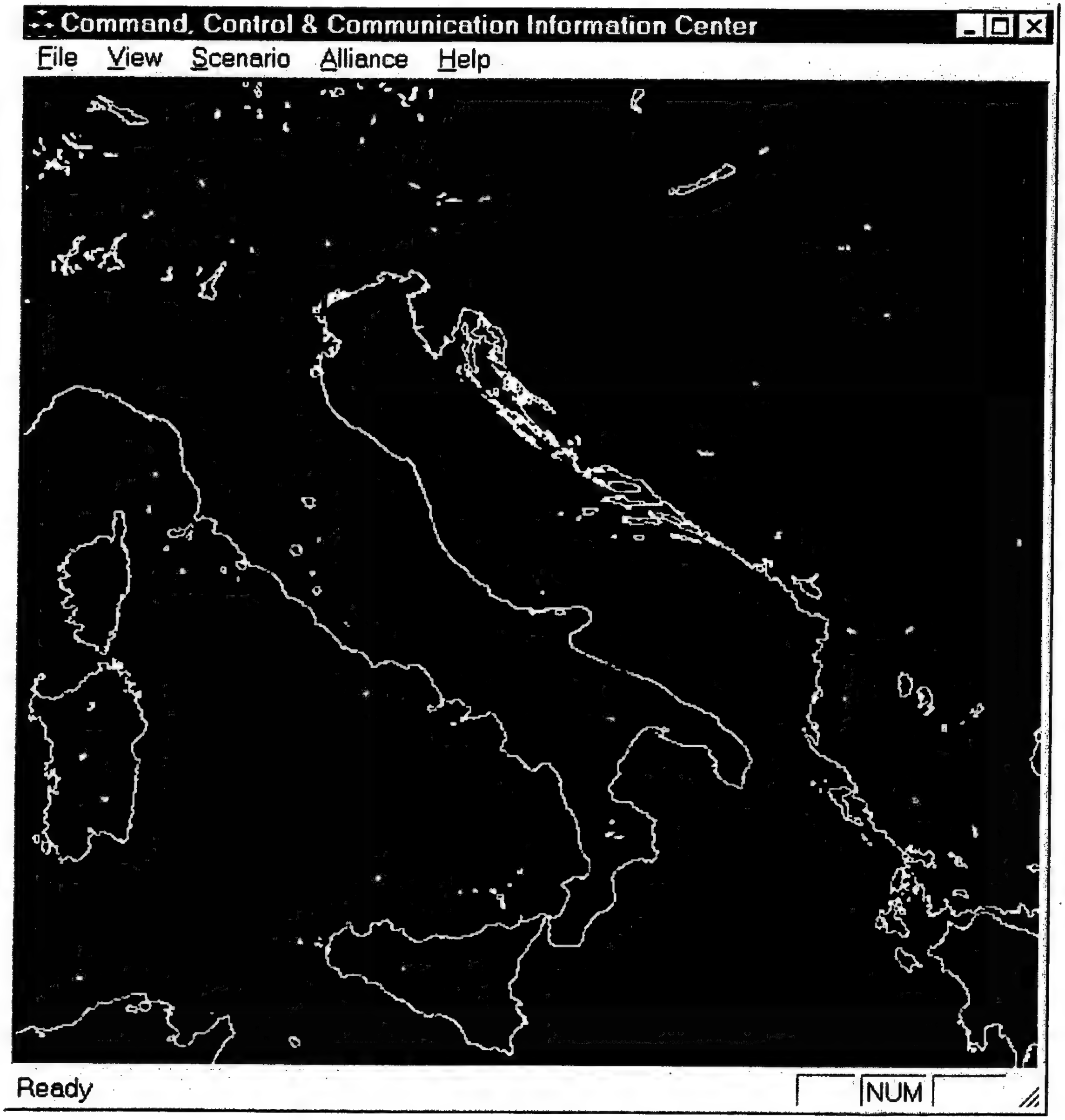


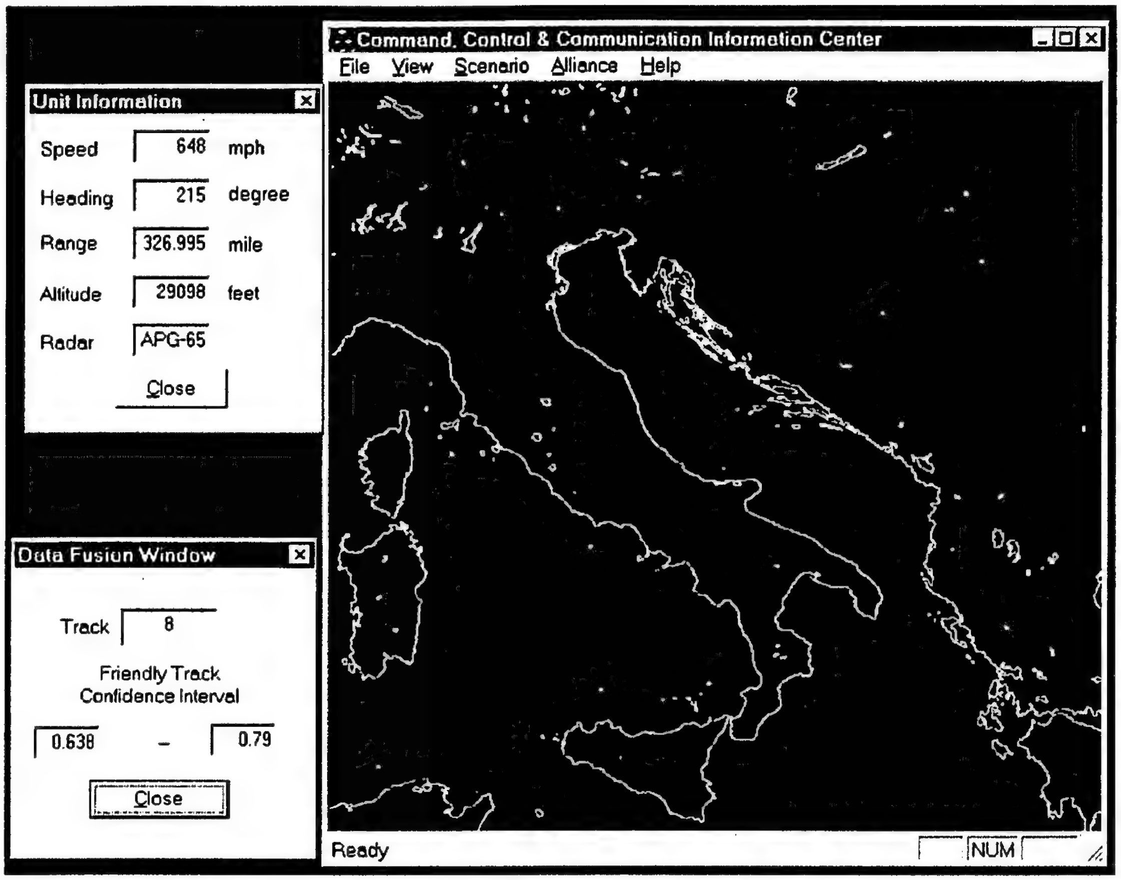
Figure 4.1. Map based radar display.

The assumed track identity can be changed by selected choices from the Alliance menu. Aircraft turned either red (hostile) or blue (friendly) after the identification, depending on the identification that participants made.

The following windows were displayed by the computer system:

1. Radar Display window

The Radar Display Window (Figure 4.1) showed aircraft as they moved in the vicinity of the participant's ship. The aircraft were shown in the window as small pink squares before they were identified. Aircraft identified as hostile turned to a red chevron, and aircraft identified as friendly turned blue half circle. Their position was updated every



Speed 648 mph
Heading 215 degree
Range 326.995 mile
Altitude 29098 feet
Radar APG-65

Track 8
Friendly Track Confidence Interval
0.638 - 0.79

Figure 4.2. Data Fusion Window and Information Window.

two seconds according to their speed and direction. The identification command, "Alliance" in the menu in Figure 4.2, consisted of two choices: Hostile Air, and Friendly Air. Clicking Hostile or Friendly caused the currently selected aircraft to be identified as hostile or friendly, respectively.

To create aircraft for each session, a pool of aircraft was created with the pre-determined Bayesian probability for each cue described above. Aircraft from the pool were then randomly divided into six scenarios (See Appendix C for details). The aircraft's initial positions were random within the limit of the Radar Display Window.

Participants could select aircraft by clicking on them using the mouse. When an aircraft was selected, it was outlined in green.

The File menu allowed the experimenters to load the map and the scenarios. The scenario menu allows the experimenter to begin or interrupt the simulation. Also, by clicking the right mouse button, the screen will zoom in to watch tracks closely. It is designed to zoom in only ten times maximum from the initial screen. Therefore, each zooming in magnifies the map three times from the initial screen. The view menu is used to return back to the initial screen which is showed in Figure 4.2.

2. Information Window

There were two selections in the "View" menu (upper left window in Figure 4.2): Track Info Window, and Data Fusion Window. Clicking the Track Info Window opened the Information Windows (labeled Unit Information in Figure 4.2) for the currently selected aircraft. The characteristics were provided in a fixed order (speed, heading, range, altitude, and radar) over the scenarios as well as participants.

3. Data Fusion Window

Clicking the Data Fusion Window, on the other hand, opened the Data Fusion Window at the lower bottom of the screen. The Data Fusion window displayed a confidence interval indicating a presumed identity for the aircraft.

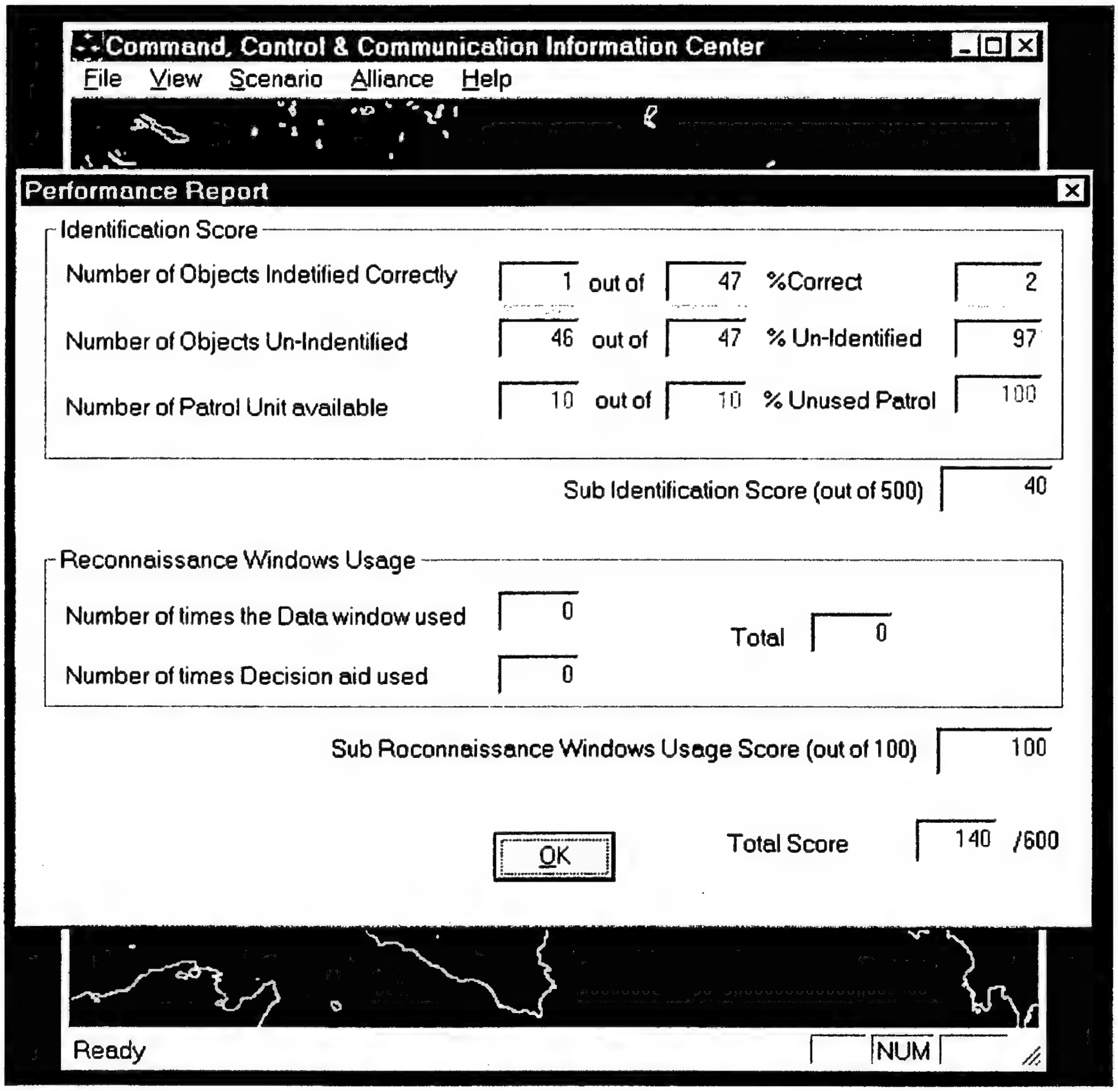
4. Performance Feedback Window

The Task Performance Feedback Window, shown below in Figure 4.3, provided feedback to participants about their performance in the scenario after the session was completed. The window consisted of two general performance measures: Identification Score and Reconnaissance Windows Usage. The total number of aircraft that were shown in the session, the number of correct identification, and percent of correct identification based on the two numbers were shown in the Identification Score. The same measures were shown for the unidentified aircraft. The total Identification Score was calculated by multiplying the percent correct by the total possible overall Identification Score (500).

The second part, Reconnaissance Windows Usage Score, was to show the number of times that participants used each window. These numbers were recorded during the simulation and were shown in the performance feedback window. Each time participants requested either information window, they lost one point. This was included to reflect the idea of relative cost to obtain information. The Reconnaissance Windows Usage score was scaled by 100 possible points.

Finally, the total score, the sum of both scores, was shown at the bottom of the window.

The simulation reads scenario files, which specify the speed, heading, and altitude of the tracks, as well as whether the track is emitting radar, and the data fusion estimate that the track is friendly. Scenario files used in the pilot experiments are included in Appendix C. Changes to any track parameters, and the time of the change, are noted in the scenario files, and stored in an event calendar as the simulation is running. At every update (every 2 seconds), a new track position is computed based on the current parameters, and the display is updated.



Performance Report

Identification Score

Number of Objects Indetified Correctly	1	out of	47	%Correct	2
Number of Objects Un-Indentified	46	out of	47	% Un-Identified	97
Number of Patrol Unit available	10	out of	10	% Unused Patrol	100

Sub Identification Score (out of 500) 40

Reconnaissance Windows Usage

Number of times the Data window used 0

Number of times Decision aid used 0

Total 0

Sub Roconnaissance Windows Usage Score (out of 100) 100

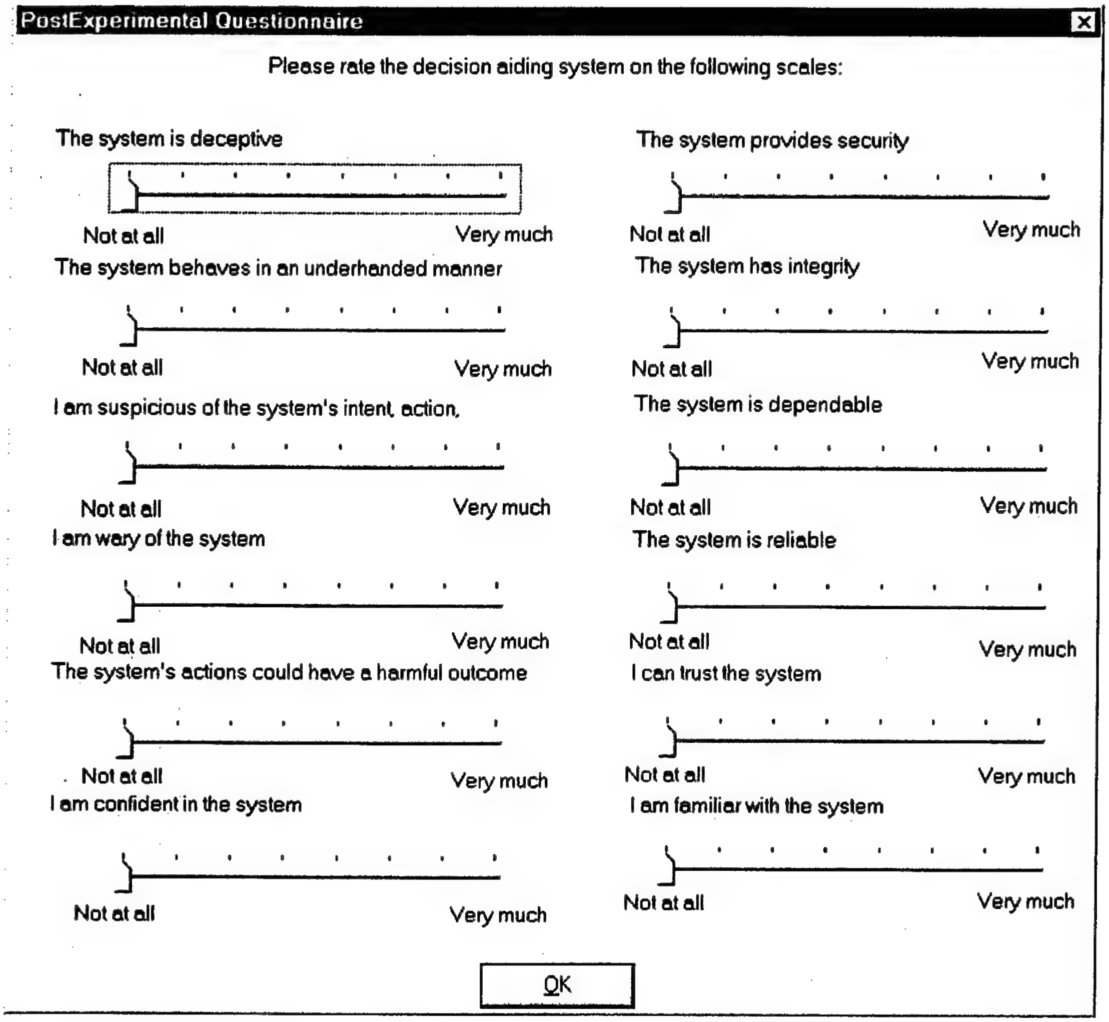
OK

Total Score 140 /600

Figure 4.3. Performance feedback window.

Finally, the 12-item trust scale described in Chapter 2 was implemented as a computer survey, as shown in Figure 4.4.

Further details on the experimental environment can be found in Appendix E, which describes the files, objects, and methods which implement the environment.



PostExperimental Questionnaire

Please rate the decision aiding system on the following scales:

The system is deceptive

Not at all — Very much

The system behaves in an underhanded manner

Not at all — Very much

I am suspicious of the system's intent, action,

Not at all — Very much

I am wary of the system

Not at all — Very much

The system's actions could have a harmful outcome

Not at all — Very much

I am confident in the system

Not at all — Very much

The system provides security

Not at all — Very much

The system has integrity

Not at all — Very much

The system is dependable

Not at all — Very much

The system is reliable

Not at all — Very much

I can trust the system

Not at all — Very much

I am familiar with the system

Not at all — Very much

Figure 4.4. Computerized trust survey.

CHAPTER 5

PILOT EXPERIMENTS IN TRUST IN AADM

5.1 Introduction

Given the theoretical framework described in Chapter 2, and the experimental environment described in the previous chapter, an initial pilot experiment was developed to test research questions in the area of human trust in an AADM environment. In particular, one factor was selected for initial study: apparent causes of failure.

5.2 Research Questions

The research question in this particular pilot experiment was to assess the impact of the cause of the failure (sabotage, hardware/software failure, or unspecified) on decision performance, and selection of information to use (i.e., implying trust in the information) over the experimental sessions. In terms of Figure 2.2, this pilot study investigated difference in trust between sabotage and non-intentional causal factors, at the automation component level. The type of malfunction was a degradation, and the surface-depth dimension was at the depth level.

5.3 Method

5.3.1 Participants

Participants were recruited from the university community and were paid $6.50 per hour for their participation. There were 10 participants in each condition, for a total of 30 participants including the control condition where participants were not given any information about either possible hardware and/or software failures or potential sabotage by enemies. Among them, 20 participants were male. There were 23 students who had taken one or more probability-related course. The average age was 26.05 years.

5.3.2 Apparatus

Experiments were run using the simulation software described in the previous chapter. Personal computers equipped with Windows NT and a Pentium 300 MHz processor were used to run the simulation. The simulation was displayed on 17" high resolution, color monitors, and participants interacted with the simulation using a keyboard and mouse.

5.3.3 Independent Variables

As noted above, the primary independent variables included in the pilot experiment was failure cause. Training materials provided to the participants differed in the description provided about the potential failure of the data fusion system. Training materials for the control condition did not mention the possibility of a failure. Table 5.1

shows the relevant portions of the training materials (see Appendix E for complete experimental materials).

Table 5.1. Portion of Task Instructions Indicating Failure Condition

Non-Intentional Failure Condition	Sabotage Condition
To perform your task, you are given a variety of electronic information sources, including an automated decision aid, which provides you with a probabilistic estimate of an aircraft's identity. This system is based on a variety of electronic and intelligence sources of information, and is combined using advanced mathematical techniques. Past experience has shown that this aid helps commanders make identification decisions, but may be subject to occasional hardware or software problems which may cause the aid to produce unreliable estimates.	To perform your task, you are given a variety of electronic information sources, including an automated decision aid, which provides you with a probabilistic estimate of an aircraft's identity. This system is based on a variety of electronic and intelligence sources of information, and is combined using advanced mathematical techniques. Past experience has shown that this aid helps commanders make identification decisions, but may be subject to intentional interference with the computer system by enemy forces which may cause the aid to produce unreliable estimates.

5.3.4.Data Collection

All data regarding participants' actions were automatically recorded by the computer simulation. Button clicks and menu selections were recorded, along with the relevant track and parameters. On-line questionnaires were used to administer the trust questions.

5.3.5. Experimental Task

During the experiment, participants clicked on unknown contacts, requested either the information window or decision aid (data fusion) window to obtain information, and made identification of the contact as hostile or friendly. Participants performed the task for six, 20 minute scenarios. In each scenario, there were between 37 and 50 contacts to identify. During the first two scenarios, the decision-aid provided accurate confidence intervals, based on the actual probability that a contact with a particular pattern of altitude, speed, and radar was friendly. The confidence interval was computed based on this probability, plus and minus randomly generated errors that ranged from 0 to 5% (two errors were generated so that the true probability did not always fall at the center of the interval). During the first 10 minutes of the third scenario, an error was introduced into the data fusion aid confidence interval. This error was computed by either shifting the intervals to one direction 5% or widening each direction 2.5%. Ten minutes after the onset of the simulation for scenario 3, the participant was informed that an error in the decision aid had been detected and corrected. Finally, three scenarios without errors were completed after the error scenario (see Figure 5.1).

The task was identical for all three conditions with the exception that in the control condition, the participant was not notified that an error was detected.

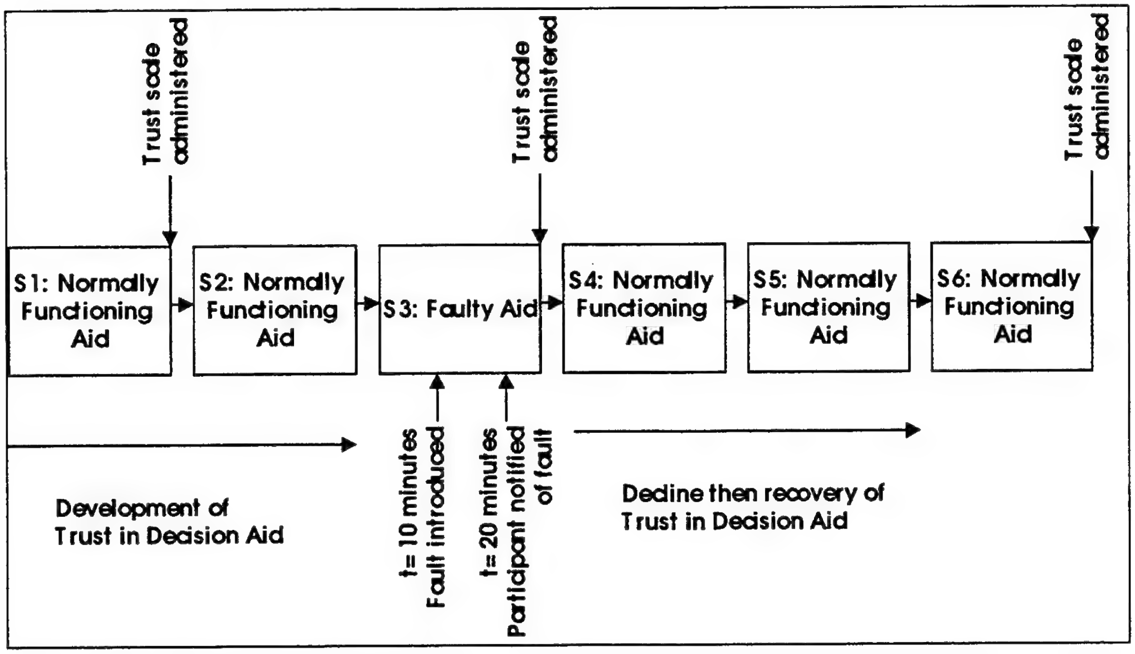

Trust scale administered
Trust scale administered
Trust scale administered
S1: Normally Functioning Aid
S2: Normally Functioning Aid
S3: Faulty Aid
S4: Normally Functioning Aid
S5: Normally Functioning Aid
S6: Normally Functioning Aid
t= 10 minutes Fault introduced
t= 20 minutes Participant notified of fault
Development of Trust in Decision Aid
Decline then recovery of Trust in Decision Aid

Figure 5.1. Organization of scenarios.

Throughout the scenario, contact parameters varied according to the following: Six different radar signatures (Lambert, 1994) were assigned to the tracks based on the pre-defined probabilities. Each track had only one radar signature; however, signatures were not completely diagnostic of hostile or friendly aircraft. The radar signature also was either available, or unavailable (that is, participants could not always access this information). Speed and altitude were also varied. To reflect the fact that these profiles can be varied over certain ranges, three overlapping categories for speed and altitude were configured based on pre-defined probabilities. These categories were overlapped with one another to represent the real-world and to prevent the participants from utilizing this characteristic as a single cue to make a decision.

In order to investigate the impact of trust, particularly after the decision aid error, on use of the information sources, participants were given a limited number of times they could access both the information window and the data fusion window. Both types of windows disappeared from the screen five seconds after they were requested. The locations of both windows were fixed in position regardless of the order that participants clicked the menu so that they could expect to see information requested at the same location. Participants could access these windows a total of 1.5 x the number of contacts in the scenario (That is, they could not access both windows for all contacts). Participants' scores reflected this tradeoff, as well as their correctness in identifying contacts, and was calculated as shown in Table 5.2. Participants were informed of the scoring method, and were shown their score after each scenario.

Table 5.2. Scoring Method Taken From Task Instructions

Due to limited resources, the best decisions will be those that minimize the time consuming requests for information. However, it is imperative to identify aircraft correctly, so that friendly forces are not inadvertently attacked, and so that friendly assets can be protected. Therefore, your score for each session will be based on both your correct identifications, and the amount of information you request to make identifications. Specifically, you will be penalized 10 points for every object you misidentified. Total of 600 points, the correct identification contributes 500 points. You also will be penalized 1 point for every time you open up the decision aid windows. You will be able to see your score at the end of each session.

Participants were given task instructions to read, and given a 10 minute practice scenario during which time the experimenter showed the participant the mouse and menu functions.

Dependent variables included the correctness of participants' actions identifying unknown contacts, the number of times participants accessed either the decision aid (data fusion) window or the information window, the score shown to participants, and the subjective trust ratings collected via the computerized questionnaire. The Trust questionnaires were given after scenarios 1, 3, and 6, as shown in Figure 5.1.

5.3.7 Experimental Design

Failure cause (3 levels) was a between-subjects variable, while scenario was a within-subjects factor. For the pilot study, 10 participants were included in each cell, and performed the experiment for all six scenarios.

		Constant Error
Failure Cause	Intentional Disruption	10 participants Scenarios 1-6
	Non-intentional Disruption	10 participants Scenarios 1-6
	No Information about the Disruption	10 participants Scenarios 1-6

Figure 5.2. Experimental design.

5.4 Results

5.4.1 Performance Measures

Performance Outcome Measure

Dependent outcome measures were derived to describe overall performance on the aircraft identification task, including participants' accuracy and use of both aid windows in completing the task. The measures used data extracted from the log files captured during task performance. The performance outcome measures that were derived describe participants' success at making correct identifications and their use of windows to make decisions.

Participants' success in identifying aircraft was described by a count of the number of aircraft in each session that a participant identified correctly. Recall the fact that the total number of aircraft that were assigned to each session was random. Furthermore, participants were instructed that they could change or identify the same aircraft more than once. This measure is a raw performance score, which could be affected by the number of aircraft randomly assigned to each session. As such, it is not an ideal measure of participants' success identifying aircraft. Another measure, percent correct was derived to address the problem. The percent correct measure is a ratio of the number if aircraft identified correctly on the attempts in a session, and the total number of aircraft identified in a session. Thus, it provides a measure of identification success based only on the number of aircraft participants' identified or the total number of aircraft assigned to each session, and factors out, for example, the fact that the total number of aircraft for those sessions were different.

Result based on the performance outcome measure is described by environmental conditions (Hardware/software failure, Sabotage, Control).

An analysis of variance was used to investigate difference in performance outcome measure between environmental conditions. Participants' environmental conditions were treated as a between-subject, fixed factor, and session was treated as a within-subject, fixed factor. Subject, nested within environmental condition was treated as a random factor.

Fault condition was marginally significant, F (2, 27) = 2.634, p = .09; post-hoc tests indicated that there was a significant difference between the control and hardware-software fault conditions, p = 0.019, and between the hardware-software failure and sabotage condition, p = 0.008, in the first session. Averaged across failure conditions, mean percent correct ranged from 61% (session 2) to 85% (session 3), with an overall mean of 74%. In Figure 5.3, this is indicated by the substantial differences in the first scenario.

Tests of within-subject factors reveals that there are significant differences among scenarios, p < 0.001, and in the interaction between scenario and instruction, p =0.046.

There was no significant environmental condition effect other than at the first scenario even though performance was somewhat higher for the sabotage condition than other conditions throughout six scenarios, as shown in Figure 5.3.

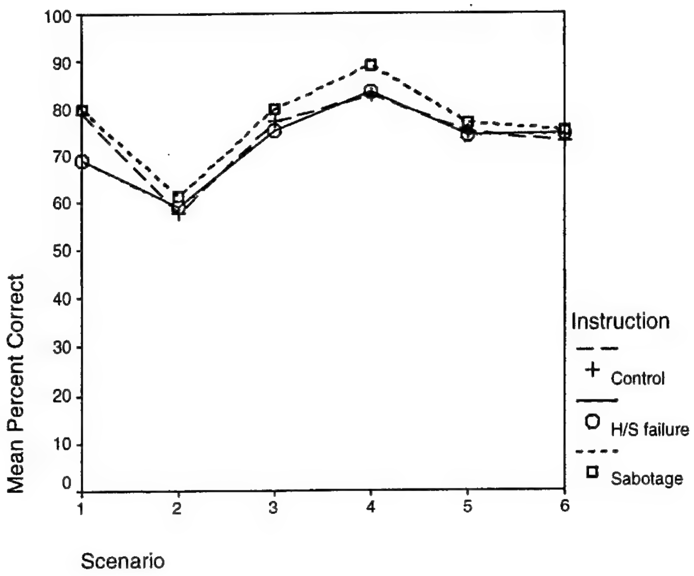


Figure 5.3. Percent correct identifications.

5.4.2 Use of Decision Aid Windows

Mean Percent Use of Either Aid

As shown in Figure 5.4, use of both decision aids windows was analyzed across the three conditions. Considering the fact that participants' had control over which decision aid/window they wanted to look up, the percent of each window could be a measure for which one they preferred over the other window. The mean percent TIW (Track Information Window) use was calculated as the number of times where participants requested the TIW divided by the total number of times either window the Data Fusion Window (DFW) or TIW was requested. Since the mean percent TIW use and the mean percent DFW use sum to 1, only the TIW use is shown. There was a significant session effect, F (5, 135) = 2.487, p = .034. From Figure 5.4, use of the information window tended to decrease from sessions 3 to 5, then recover in the final session. Inspection of Figure 5.4 also suggests that participants in the sabotage condition tended to use the information window less, and thus the decision aid window more, although the difference between fault conditions was not statistically significant, F (2, 27) = 1.213, p = .313. This may indicate that participants were attributing the failure to the track information window, rather than the data fusion window as expected.

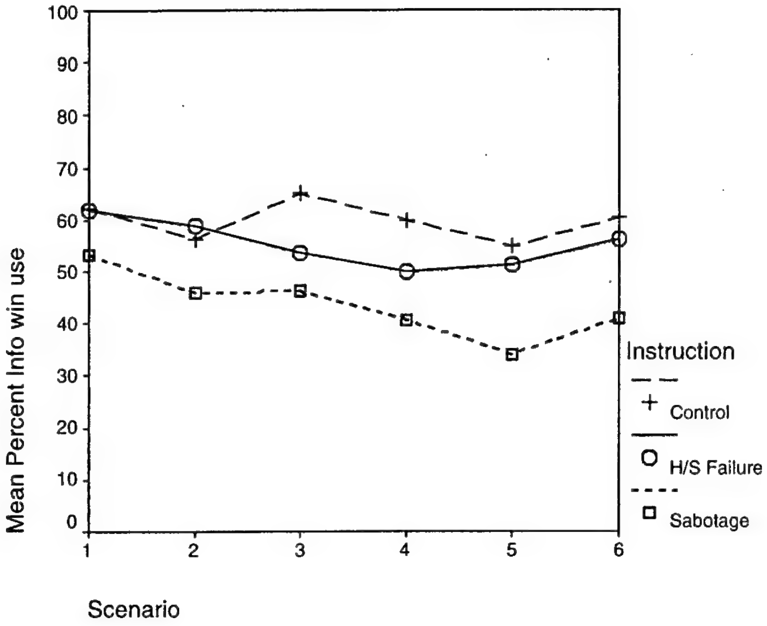


Figure 5.4. Mean percent use of the information window.

Use of both windows for each track

Participants were instructed that they were capable of requesting data for the selected track more than once for each track, but not more than 1.5 times the total number of tracks. Mean percent of times that participants requested for both windows was analyzed, and is shown in Figure 5.5. Though there was no significant difference between condition, $F(2, 27) = 0.498$, $p = .613$, there was a significant session effect, $F(5, 135 = 2.381$, $p = .042$. Participants tended to be less likely to select both windows, across

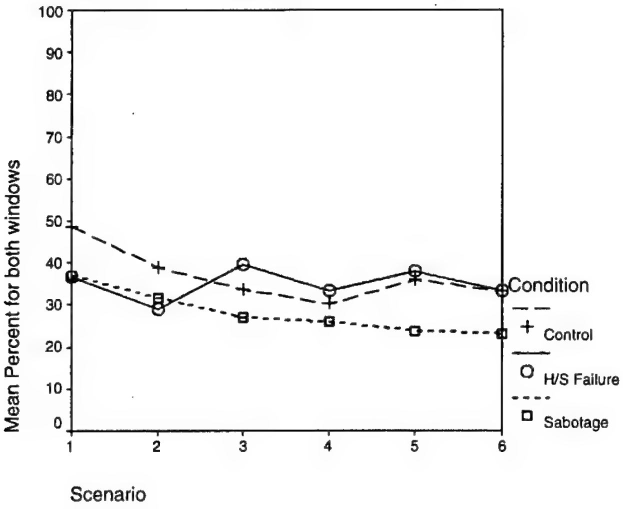


Figure 5.5. Mean percent use of both windows.

the scenarios. Also, there appears to be a non significant trend for participants in the hardware/software failure condition to increase their use of both windows during and after the third scenario, in which the failure occurred.

Use of either window

Finally, participants' use of just one of the two windows was analyzed. After removing the cases of requesting both windows for each selected track, mean percent use of either window was analyzed to investigate the participants' behavior when they had only one chance to request for information. While the main effect of condition on Mean Percent information window use was insignificant, F (2, 27) = 1.896, p = 0.170, post-hoc analysis showed that there was marginally significant difference between the control and sabotage conditions over the six trials, p = 0.071. Information window only use, shown in Figure 5.6, was lower in the sabotage condition (hardware/software failure was in between). The mean percent of data fusion window only use showed no significant differences among the environments. It is interesting to note that the sole use of the information window use declined after the error was identified to participants in the third scenario, but then tended to recover. Again, this result suggests that participants were interpreting the failure as being in the task information, rather than the data fusion window.

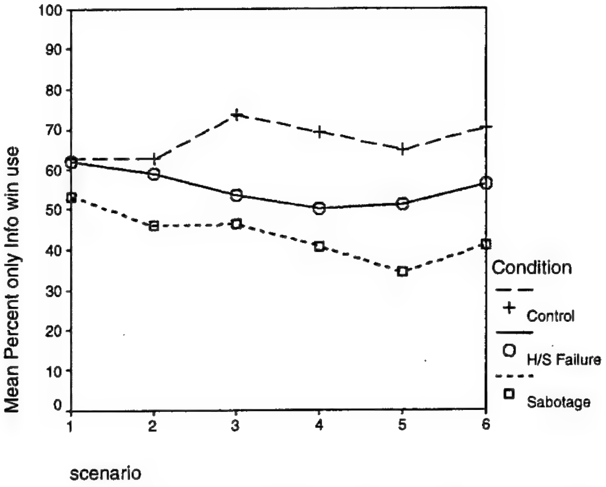


Figure 5.6. Mean percent use of the information window only.

Measures against total number of tracks

Use of information sources in a session, per total number of contacts in a session, was also investigated. These measures scale the measures of information use by the number of contacts to be investigated and identified. The use of the information window and the decision aid window were analyzed. While there was no significant effect of fault condition on the use of the decision aid window, there was a significant impact of fault condition on information window use, $F(2, 27) = 2.943$, $p = .07$. Inspection of Figure 5.7 indicates that participants in the sabotage condition seemed least likely to select the

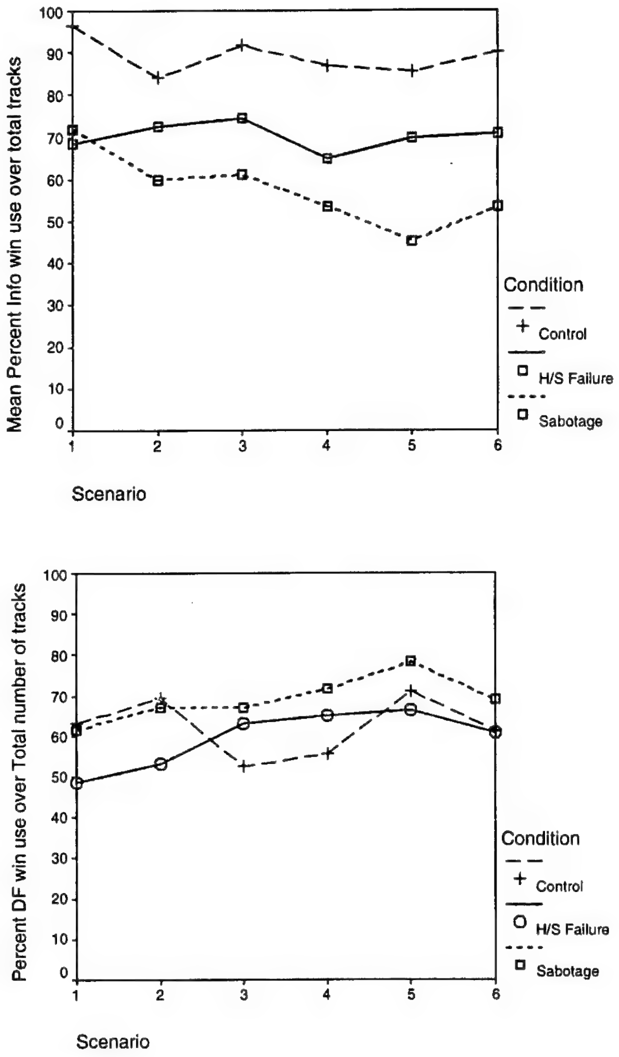


Figure 5.7. Percent of data fusion window and information window scaled by the number of tracks per scenario.

information window on a track-by-track basis, and participants in the control condition the most likely, while across the three conditions there was a trend toward decreased information window use after the third session.

Inspection of cases where the use of only one window was selected shows a similar pattern, though fault conditions were not significantly different (See Figure 5.8). Use of the decision aid window changed over sessions, $F(5, 135) = 3.852, p = .003$, tending to increase after the third session and decreasing from the fifth to sixth session. In general, across the information use measures, there appears to be a trend for participants in the sabotage condition to make less use of the information window, and for participants across conditions to make less use of the information window and more use of the decision aid window after the third session, perhaps indicating that participants attributed the failure to the information window rather than the decision aid window.

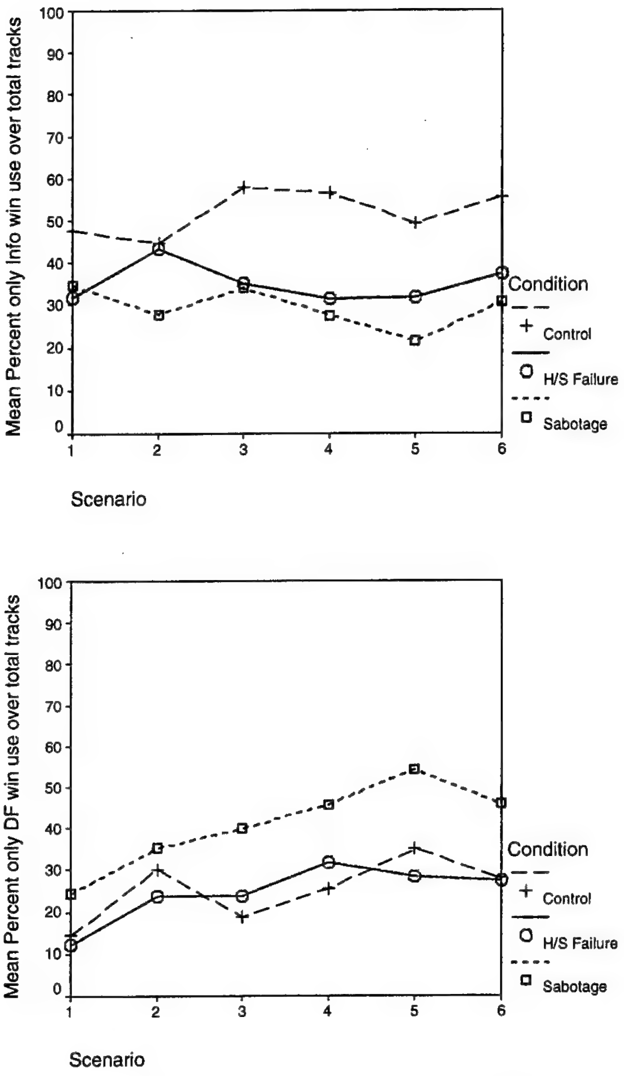


Figure 5.8. Use of each information source only, by number of contacts per scenario.

5.4.3 Trust Questionnaire

The questionnaire described in Chapter 2 was given to participants in the electronic form three times (after trials 1, 3, and 6) throughout the experiments. Participants rated their agreement with the statements on a continuous scale measured from 0 to 7 with endpoints labeled "Not at All" and "Very Much."

Statistical analyses were performed on the participant's responses to the questionnaire. During the analyses, question and scenario were treated as within-subjects factors, while fault condition (hardware-software, sabotage, and control) was treated as a between-subjects factors. Questions are shown in Table 5.3

Table 5.3. Questions in the Post-scenario Questionnaire. Positively Framed Questions are in Italic, Negatively Framed Questions are in Plain Text.

The system is deceptive
The system behaves in an underhanded manner
I am suspicious of the system's intent, action, or output
I am wary of the system
The system's action will have a harmful or injurious outcome
I am confident in the system
The system provides security
The system has integrity
The system is dependable
The system is reliable
I can trust the system
I am familiar with the system

Results provide information both about the structure of the questionnaire itself, and the relationship between the questionnaire and the behavioral data summarized above.

First, analysis of the responses to all twelve questions indicated that there was a main effect of question, $F(11, 297) = 4.768$, $p <= .000$: questions did tap into different concepts.

There is also a significant question by condition interaction, $F(22, 297)$, $p = .037$. Further interpretation can be made by inspecting Figure 5.9. For the control and sabotage conditions, responses for both positively framed questions (displayed with dashed lines) tended to be similar to, and intermingled with, responses for negatively framed questions (displayed with solid lines). For the hardware/software failure condition, negative responses tended to be clustered together, and higher, than positive responses.

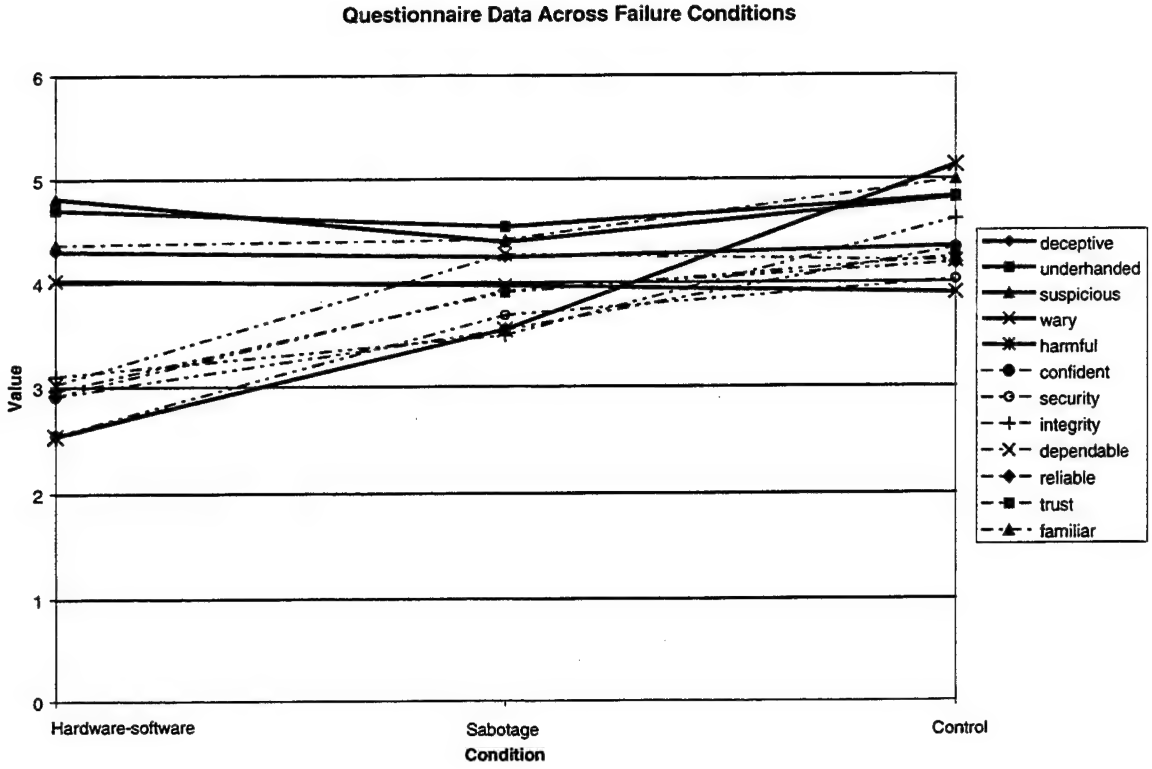


Figure 5.9. Questionnaire data across failure conditions. "Positive" trust related questions are indicated with dashed lines, while "negative" trust related questions are indicated by solid lines.

Tests were also performed on negatively and positively framed questions, separately. For negatively framed questions, there was a significant difference between questions, *F*(4, 108 = 5.061, *p* = .011, a significant question by condition interaction, *F*(8,108) = 10.348, *p* = .033, and a significant question by condition by scenario interaction, *F*(16, 216) = 1.337, *p* = .047, again indicating that the questions in fact tapped into different concepts of trust as conditions varied. Inspection of Figure 5.10 indicates that in particular, the notion of harm appeared different across conditions, with least agreement with the statement "The system's action will have a harmful or injurious outcome" occurring in the hardware-software condition. On a scenario-by-scenario basis (see Figure 5.11), responses overall seemed to decrease in the sabotage condition after the sixth scenario. Additionally, (excluding the concept of harm, which was low throughout), responses in the hardware software condition were similar after the first session, but showed more differentiation by the sixth session, with the question regarding suspicion receiving the highest rating, at the question regarding wariness the lowest.

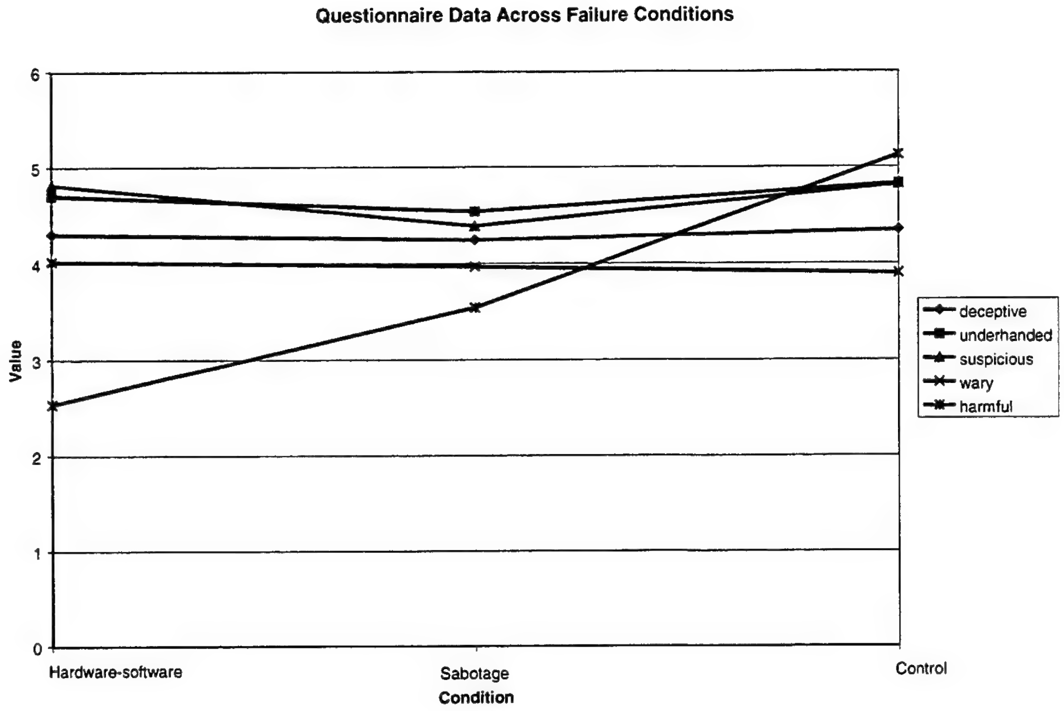


Figure 5.10. Questionnaire data across failure conditions for the negatively framed questions.

Similar tests were performed on the positively framed questions. There was a significant effect of question, *F*(6, 162) = 8.502, p < .000, significant question by scenario interaction, *F*(12, 324), *p* = .005, and significant question by condition by scenario interaction, *F*(24, 324) = 1.575, *p* < .044. These results are shown in Figures 5.12 and 5.13. From Figure 5.12, the question regarding familiarity is rated higher across all conditions. Additionally, from Figure 5.13, agreement with the question regarding familiarity tended to rise across sessions particularly for the sabotage and control conditions, while overall responses declined slightly for the sabotage condition, and rose for the control condition.

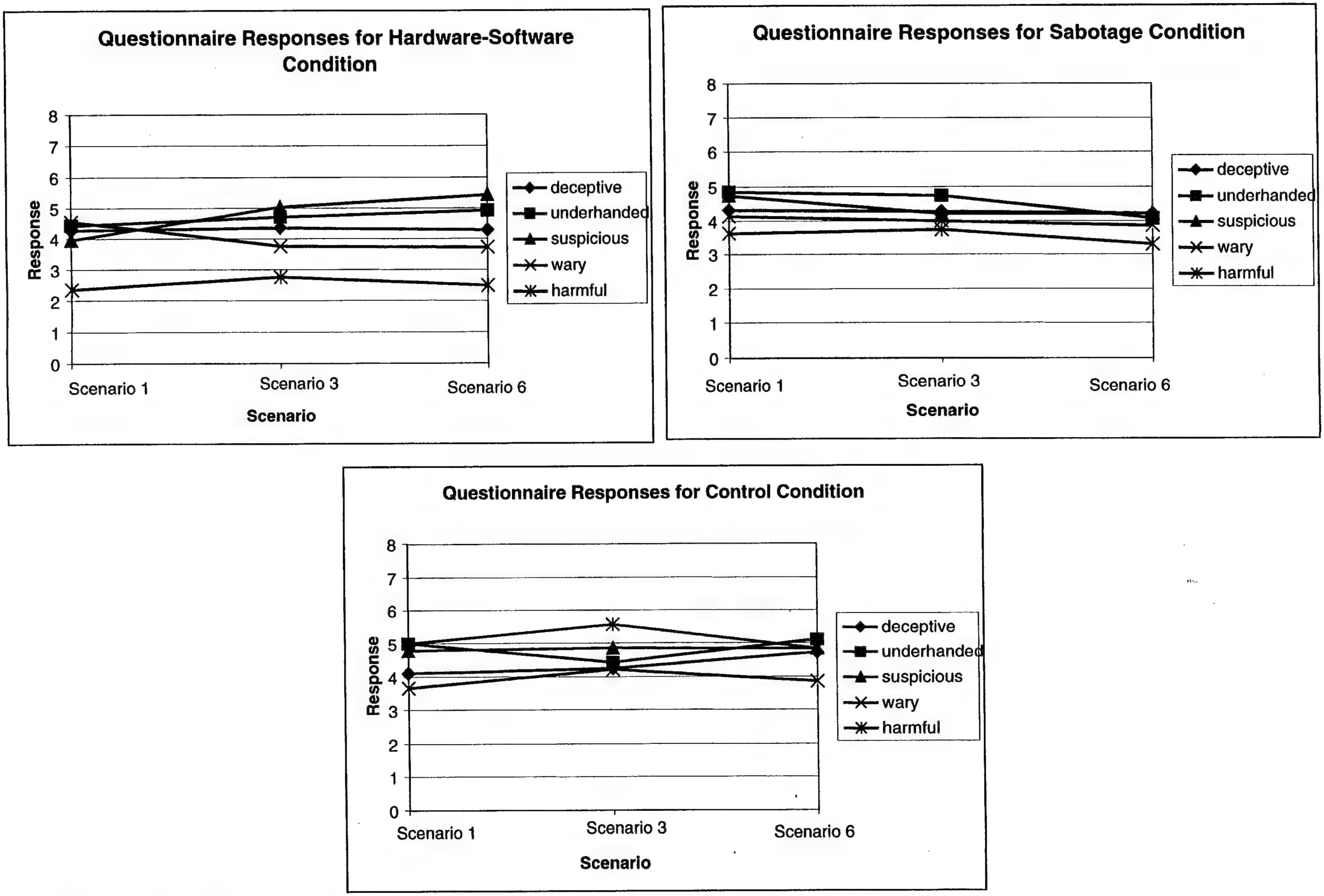


Figure 5.11. Questionnaire Responses for Negatively Framed Questions by Condition and Scenario.

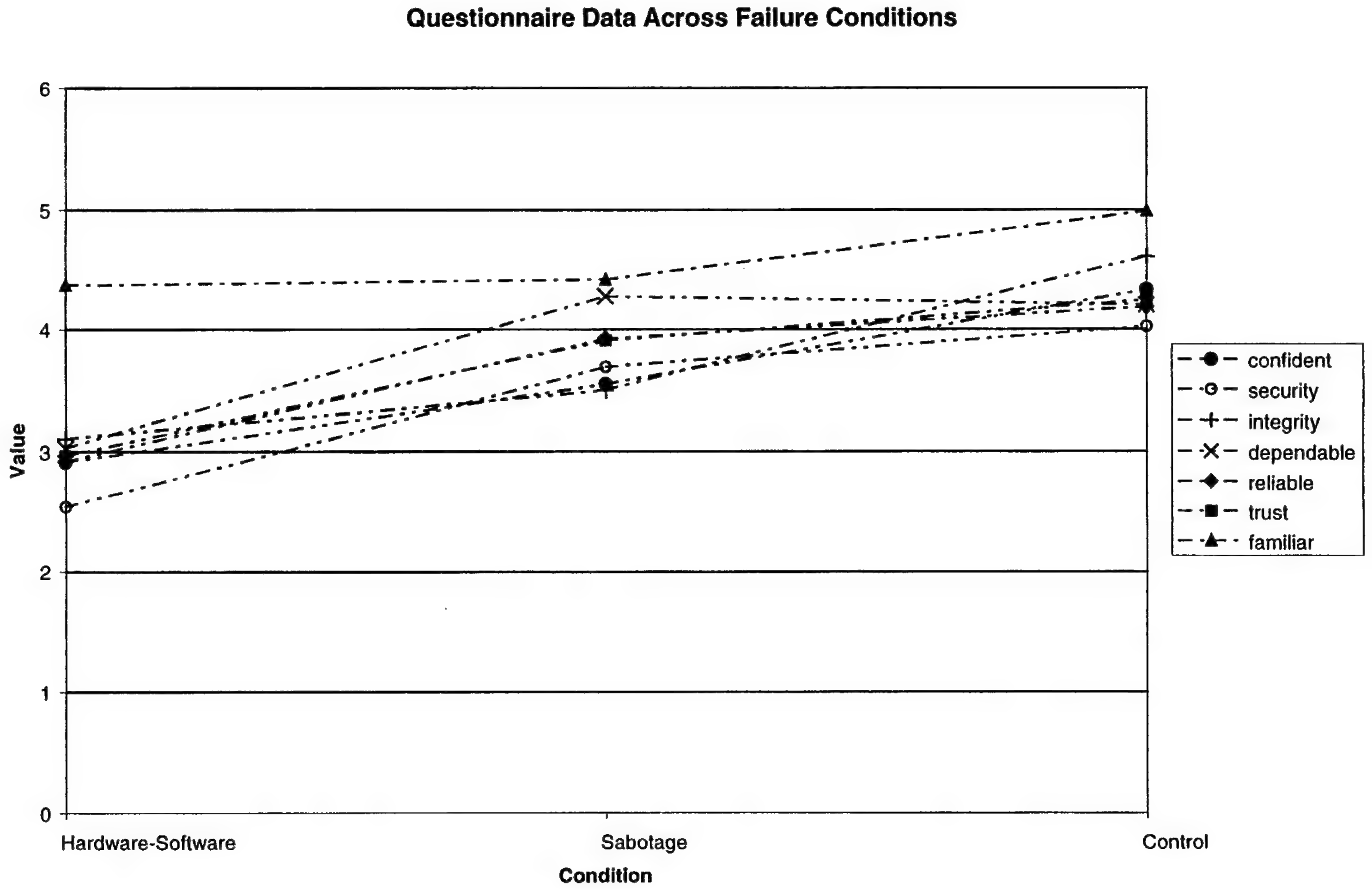


Figure 5.12. Questionnaire responses for positively framed questions by condition.

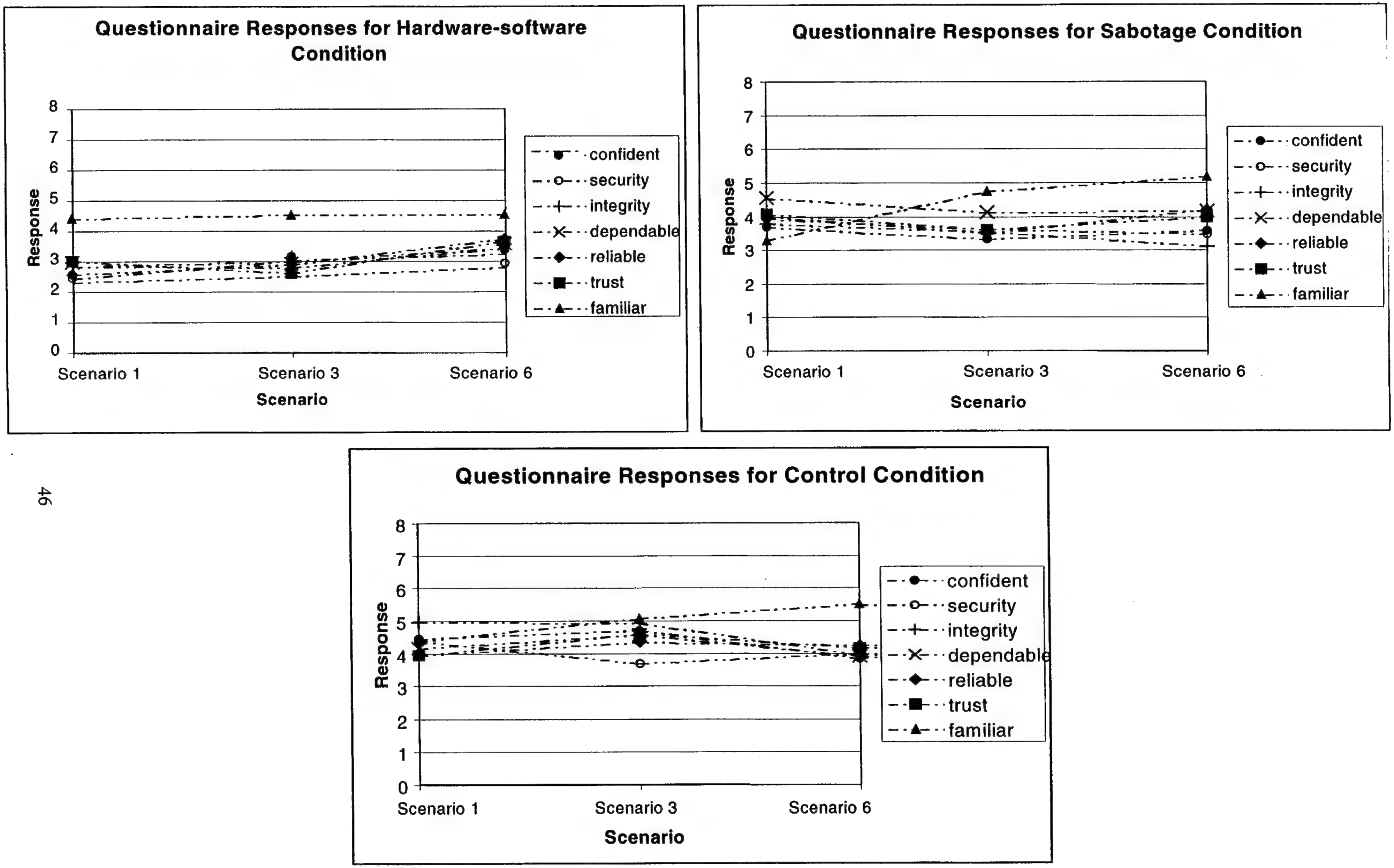


Figure 5.13. Questionnaire responses for positively framed questions by condition and scenario.

Finally, results from all positively framed questions, and all negatively framed questions, were combined, and compared to see if there was an overall difference in how participants answered the two types of questions, and thus perhaps their general view of the information systems. These results are shown in Figure 5.14. Responses to positively framed questions were significantly lower than negatively framed questions, $F(1, 27) = 2.96$; $p = .096$. There were no significant interactions with scenario or condition.

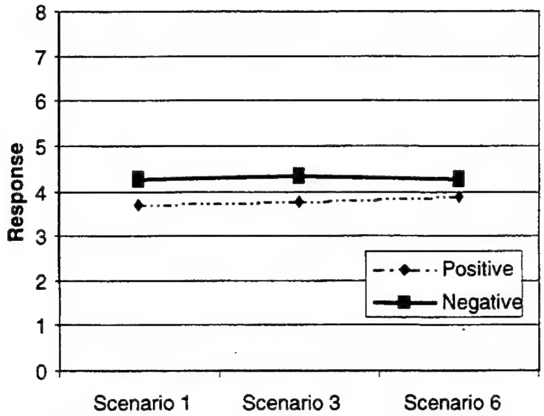


Figure 5.14. Responses for positively and negatively framed questions.

5.5 Discussion

In general, results from the experiment indicated that participants showed some tendency to reduce their use of one of the two information sources (the information window) after a fault occurred, and that participants in the sabotage condition tended to be less likely to use the information window. These results suggest that contrary to prior expectations, participants in the sabotage condition were not more suspicious of the data fusion information, suggesting possibly that participants attributed the potential for faults to the information window and not the decision aid. Future experiments could clarify this point by expanding the training and explanations provided to participants, as well as the magnitude and description of the errors in the aid.

Overall, the results are interesting, for several reasons. First, given the tendency for some difference in aid utilization between fault conditions, the inclusion of this dimension - causal factors - in the theoretical framework described in Chapter 2, is supported. Future work in this area, both in terms of laboratory experimentation, and the development of displays and aids to support AADM, should continue to consider such a dimension. For example, any automated algorithms to detect potential faults should attempt to identify and display the type of fault with respect to sabotage or natural failure conditions. Second, the tendency for a pattern of less use, then some recovery by the last session is consistent with prior work on dynamic patterns of trust, suggesting both that sabotage of a decision system may induce disuse for a period of time but not permanently induce distrust. On the one hand, this might indicate that repeated attacks would be necessary to sustain a level of disuse in an adversary's system, while on the other, it indicates that if it is possible to bring the speed at which friendly decision-makers' regain trust in a system after an attack, it may be possible to mitigate the effects of such attacks. Although results presented here were consistent with prior studies, they were of limited magnitude. Further work is needed to more definitively determine the types of failures, and the magnitudes of such failures, that could cause more debilitating effects.

Finally, results from the trust questionnaire indicate that participants responded differently to different items in the trust questionnaire, and that though they had overall higher negative beliefs about the system than positive beliefs, there was not a clear differentiation in responses between responses to positive and negative scale items. These results support those found in developing the trust questionnaire – particularly, the fact that there could be different concepts within trust, and that both positively and negatively framed concepts could be associated with trust - and thus provide some validation for the questionnaire items. Additionally, there were differences in responses to the questionnaire based on fault condition, suggesting both that the questionnaire was sensitive to differences in task conditions, and also that the different levels of fault conditions were in fact affecting participants' trust.

Overall, results indicated that as suggested by the proposed theoretical framework, differences in fault causation may impact operators' trust in, and use of, information systems, and that the trust questionnaire could identify differences in concepts of trust as conditions varied.

CHAPTER 6

IMPACT OF GRAPHICAL DISPLAY FORMAT ON DECISION-MAKING UNDER UNCERTAINTY

6.1 Introduction

In addition to trust, another factor which may influence the utility of data fusion based decision aids, and the influence of these aids on the decision making process, is the form in which the uncertain information determined by these aids is presented to decision-makers. Uncertain or probabilistic information can be shown in a variety of formats ranging from simply text to graphical representations to text/graphical hybrids. Past research has focused on representing position, direction and identity uncertainty in a format that reveals the true probabilistic nature behind the data (Andre and Cutler, 1998; Banbury, Selcon, Endsley, Gorton, and Tatlock, 1998; Kirschenbaum and Arruda, 1994).

Position uncertainty deals with how to represent the possible places an object may inhabit. Environments in which this type of uncertainty plays an important role include commercial aviation and military sonar/radar. Andre and Cutler (1998) investigated this form of uncertainty with the use of a task in which a pilot would have to play "Chicken" with a circular object, they called a meteor. The pilot's goal was to come as close as possible to the meteor without collision. To represent the position uncertainty, a circular ring surrounded the meteor. The ring varied in size dependent upon uncertainty level. Collision frequency was found to be far less when the ring was displayed: without the ring, participants appeared to dismiss the fact that uncertainty was present in the system. Kirschenbaum and Arruda (1994) conducted a similar experiment which investigated the effect of different displays of position uncertainty on a decision-making task as to when and where to fire at a target. Participants were shown either a graphical representation of position uncertainty in the form of an ellipse around the target or a verbal indicator that ranged from poor to fair to good. The elliptical aid was found to be superior to the verbal indicator in cases of moderate to high difficulty scenarios. Overall, it appears that the use of a visual position uncertainty aid helped the performance of the user.

Aids which present heading uncertainty attempt to display all the possible future directions an object may move. Andre and Cutler (1998) tested three different types of heading uncertainty aids in a simulated anti-aircraft task: a textual description and two graphical representations that utilized either arcs or rings. The three aids improved user performance when compared with a no aid condition. The arc-based aid, which represented the uncertainty in direction by utilizing an arc that covered the entire angle of possible movement heading, provided a slight advantage over the other two aids.

Finally, identity aids strive to give the user an idea of how accurate the identification of an object is. Currently most aids display this information in the form of probabilities. Banbury et al. (1998) investigated how the context in which information is displayed affects a decision-making task. Participants were asked to make a shoot/no-shoot decision based on a probabilistic estimate of an aircraft's identity, presented as a numeric percentage. Results showed an impact of estimate uncertainty - participants were found to have a reluctance to shoot when uncertainty was greater than 9%. Additionally, presenting a secondary target identification (e.g., not just the chance that is a hostile fighter, but also the chance that it is a friendly aircraft) also impacted decisions to shoot. Participants were more hesitant when a secondary, friendly, target identification estimate was given.

Another way in which the graphical form of information presentation could be used to represent uncertainty is through the use of degraded or distorted images. Lind, Dershowitz, Chandra, and Bussolari (1995) provide evidence that the form of displayed information may affect the use of uncertain data. In a study to investigate the extent to which the graphic depiction of weather systems could be degraded (due to technical limitations) and still be acceptable to general aviation pilots, Lind et al. found that pilots' estimates of weather hazards increased as the graphical distortion increased. In this case, the distortion took the form of larger polygon/ellipse shaped depictions of weather patterns, in contrast to the non-distorted continuous, fine-grained representation. This increase in perceived risk might indicate a decrease in subjects' confidence of their understanding of the current specific weather patterns.

Thus, there is some indication that iconic representations based on degraded or distorted images may be used to convey the uncertainty associated with a decision aid estimate. In the following pilot study, we investigated properties of distorted and blended icon sets intended to convey uncertain information about an object's identity as either potentially hostile or friendly. Future experiments will investigate the impact of a subset of these icons, selected based on the pilot study results, on a decision-making task.

6.2 Pilot Study Method

6.2.1 Participants

Twenty participants, all undergraduate students, were paid $6.00 per hour for their participation in the pilot study.

6.2.2 Experimental Design

Five sets of pictures were chosen to represent the identity of an object as either hostile or friendly. These picture sets were classified as either abstract (without an obvious associated meaning), iconic (with an associated meaning), or both. Picture pairs were chosen in order to allow for the entire spectrum from friendly to hostile to be represented. Figure 6.1 shows the pictures used in the experiment.

In order to represent the probabilistic nature of the information graphically, a series of thirteen icons were created to represent a range of probabilities (i.e., from p(Hostile) – 0.0 to p(Hostile) = 1.0). The iconic and abstract picture pairs were distorted and blended using a pixelizing function found in Adobe PhotoShop 4.0. For example, the 50% friendly/50% hostile picture blended both of the pictures in a pair together. For the colored icons, the series of icons was created by coloring each pixel in the icon as either

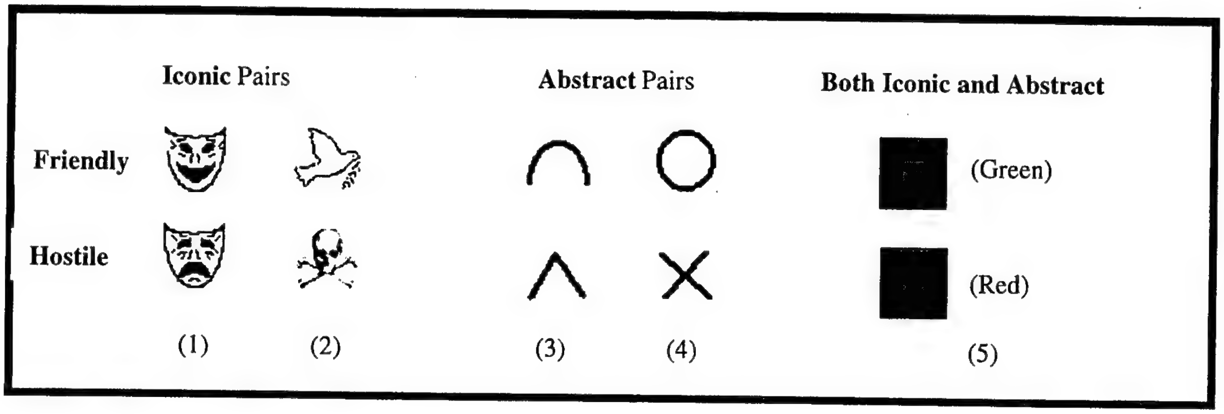


Figure 6.1. Five pairs of icons representing object identities as either hostile or friendly.

green or red based upon the probability desired. To illustrate how the pixelizing function works, the series of the distorted and blended pictures for picture pair (1) are shown in Figure 6.2.

Each participant performed a series of tasks involving all five sets of icons. Ten participants performed the tasks under a "friendly" framing condition, and ten participants performed the tasks under a "hostile" framing condition. In the friendly framing condition, participants were given task instructions which described the icons as more or less friendly. In the hostile framing condition, icons were described as more or less hostile.

6.2.3 Procedure

The three experimental tasks were designed to measure whether the icons could be correctly sorted and assigned a probability rating according to the expected probabilities that the icons represented. Participants performed each of the tasks five times: once for each icon pair (see Figure 6.1).

In the first task, a timed sorting task, participants were asked to sort cards into piles according to the icon printed on the card. Participants were asked to create piles containing the same icon. There were five instances each of the 13 possible icons in a set, for a total of 65 cards. The time to sort the cards, and sorting errors, were collected.

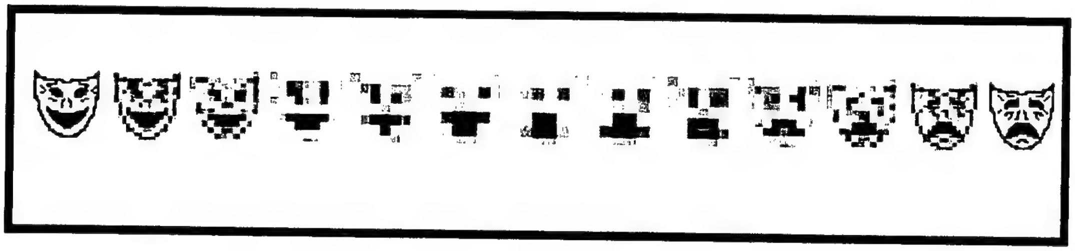


Figure 6.2. Series of 13 icons representing a range of probabilities that an object is hostile or friendly: from a probability of 100% friendly to 100% hostile.

In the second task, participants were asked to order the set of thirteen pictures from most to least friendly (or hostile), depending on the framing condition. They were not told which icons corresponded to the hostile or friendly ends of the scale (e.g., they were not told that a circle represented a most friendly, and an "x," least friendly). Participants performed this task using a Visual Basic computer program, through which they could drag and drop the icons into the desired order. The ordering of the icons was recorded automatically by the computer.

For the third task, participants were asked to rate each icon on a continuous scale, with end points of least and most friendly (or hostile). Participants marked their rating along a line connecting the endpoints; this distance was later measured and scaled based on the length of the line, and used to identify their rating.

6.3 Pilot Study Results

6.3.1 Card Sorting

The times to sort cards based on the icon printed on the card did not differ significantly across picture pairs. Thus, the relative difficulty of identifying and sorting the thirteen icons did not appear to differ across sets.

6.3.2. Ordering

The order of the thirteen icons in each icon pair set was determined for each participant, for the hostile and friendly framing conditions, resulting in ten orders per icon pair for each framing condition. These orders were used to compute an average ranking for each icon for the five pairs, for both framing conditions. Ordering these average rankings resulted in an average order for each set, for both framing conditions (a total of 10 average orders). These average orders were correlated with the expected order (based on the way the icons were created), and a Spearman correlation coefficient was computed. These coefficients are shown in Table 1. All correlations were significant at the .01 level of significance, indicating that overall, participants were able to correctly order the sets of icons according to the intended levels of uncertainty.

Table 6.1. Spearman Correlation Coefficients Comparing Average Rank Orders to Expected Order for 5 Icon Pairs.

	Framing	
Icon Pair	Friendly	Hostile
Mask(1)	1.000	0.929
Dove(2)	1.000	0.984
Inverted V-U (3)	0.934	0.984
Circle-X(4)	1.000	0.951
Color(5)	1.000	0.890

Individual participant data was also examined: Spearman correlation coefficients were computed comparing each participant's order to the expected order, for both framing conditions. These correlations are indicated in Tables 6.2 and 6.3, corresponding to the Friendly and Hostile framing conditions, respectively. Correlations in bold are *insignificant* at the .05 level of significance. Inspection of Tables 6.2 and 6.3 shows that on a participant-by-participant basis, ordering was more consistent and correct in the friendly framing condition than the hostile framing condition. Note that negative correlations simply indicate that the participant reversed the hostile and friendly ends of the scale (they were not told which icons corresponded to which endpoints before the experiment). It is interesting to note that even for the two "abstract" icons, reversals happened at a rate less than chance, indicating that perhaps there was some meaning intrinsic to the abstract icons.

Table 6.2. Individual Correlation Coefficients for Each Participant (Friendly Framing Condition; Bold Correlations are *Insignificant*).

Participant #	Mask	Dove	V-U	Circle	Color
	(1)	(2)	(3)	(4)	(5)
2	0.995	1.000	-1.000	1.000	0.995
4	0.989	0.995	1.000	1.000	1.000
6	1.000	1.000	0.995	1.000	1.000
8	1.000	1.000	-1.000	1,000	0.995
10	1.000	1.000	1.000	1.000	1.000
12	1.000	-1.000	-1.000	-1.000	-1.000
14	0.984	1.000	1.000	1.000	0.995
16	0.962	1.000	1.000	1.000	1.000
18	0.995	0.995	1.000	1.000	1.000
20	0.978	1.000	**-0.440**	**0.374**	1.000

Table 6.3. Individual Correlation Coefficients for Each Participant (Hostile Framing Condition; Bold Correlations are *Insignificant*).

Participant #	Mask	Dove	V-U	Circle	Color
	(1)	(2)	(3)	(4)	(5)
1	**0.126**	**0.115**	**0.115**	0.115	**0.115**
3	1.000	1.000	1.000	1.000	0.995
5	0.566	**-0.038**	0.544	**-0.297**	0.665
7	**0.412**	**0.093**	**0.115**	**0.148**	**0.088**
9	**0.005**	0.714	**0.099**	**0.044**	**0.181**
11	0.978	1.000	1.000	1.000	**0.434**
13	**0.148**	1.000	1.000	0.995	**0.456**
15	0.978	1.000	0.978	0.995	0.989
17	0.995	0.995	1.000	1.000	1.000
19	**-0.165**	0.516	**0.280**	**0.440**	0.835

6.3.3 Rating

From the data collected on individual picture ratings, an average rating was calculated for each picture within a picture pair category. These averages provided a range of estimates of the friendliness or hostility of each picture pair (Tables 6.4 and 6.5).

Table 6.4. Average Rating Spread for 5 Icon Pairs (Friendly Framing)

	Mask (1)	Dove (2)	V-U (3)	Circle (4)	Color (5)
High Rating	88.67	97.93	96.64	97.73	98.59
Low Rating	4.22	4.06	3.67	8.52	11.33

Note: Ratings for Dove, V_U, and Circle were corrected to account for obvious and consistent reversals between hostile and friendly endpoints (Participants were not told a priori which endpoint to assign to hostile, and which to friendly).

Table 6.5. Average Rating Spread for 5 Icon Pairs (Hostile Framing)

	Mask (1)	Dove (2)	V-U (3)	Circle (4)	Color (5)
High Rating	66.56	72.27	62.11	74.06	63.69
Low Rating	24.22	17.97	38.20	21.48	14.30

6.4 Display Study

Results from the pilot study indicated that sets of distorted icons could be appropriately ordered, and span a range of descriptive levels. A second study was conducted to determine the potential utility of these icons on a dynamic decision making task.

Seventy-two volunteers were recruited from the university population for this study, and were paid for their participation.

6.4.1 Experimental Design

Participants in the experiment were asked to identify objects as hostile or friendly in a dynamic, graphical environment (see Figure 6.3), based on probabilistic information about the object's identity. There were two primary manipulations tested in this study: picture type, and display type. Three picture pairs were chosen from those used in the pilot study: the dove-skull, modified military (Abstract pair #3 in Figure 1), and red-green color pair. Three display manipulations were also created. In the first, the degraded image only condition, icons became more or less degraded to indicate the current probability that the object was friendly. In the second condition, in addition to the icons changing in degradation level, icons were annotated with a numeric probability corresponding to the current probability the object was friendly. Finally, in the third condition, objects were indicated by a static, non-degraded icon (either the hostile or friendly endpoints) and the probability. This manipulation was intended to see if degraded images could be used as a viable alternative to numeric probabilities on displays, and if the use of such images, either alone or in combination with numeric probabilities, could cause decision-makers to behave differently (e.g., more conservatively with respect to the uncertain information) than if only numeric probabilities were available.

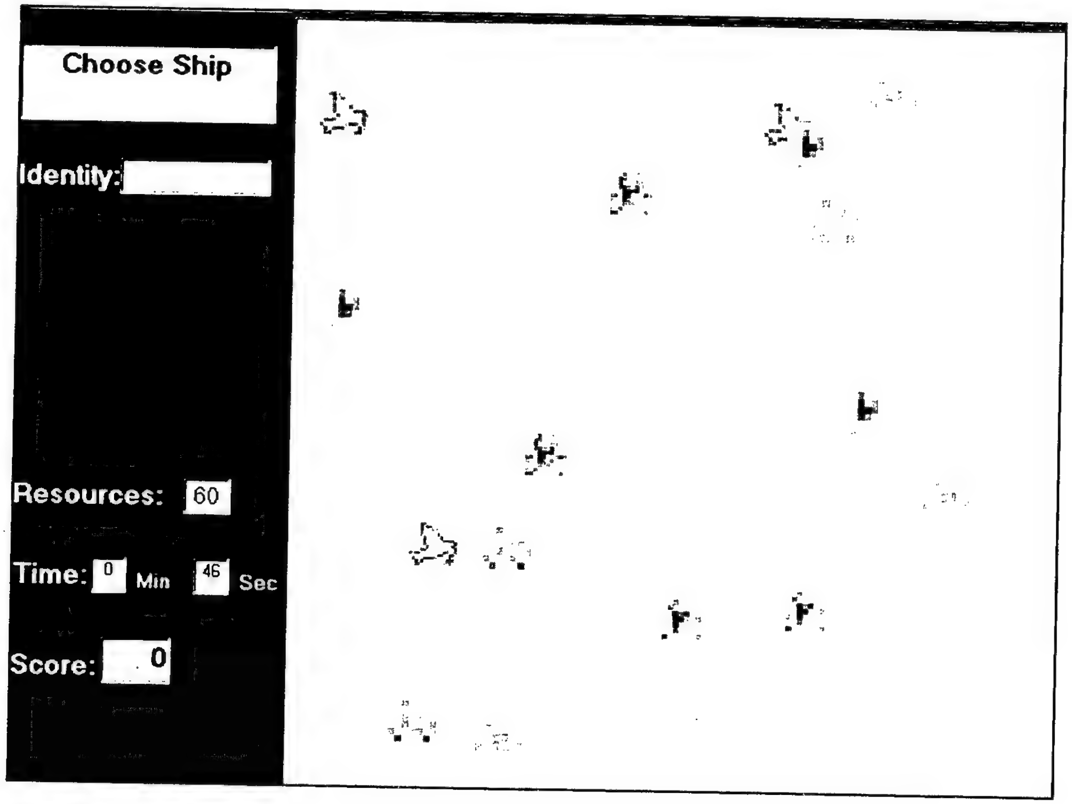


Figure 6.3. Experimental environment.

6.4.2 Experimental Task

Participants were asked to identify 40 objects per trial, for ten trials. Each object was randomly assigned an initial probability of being either friendly or hostile, and also an actual identity, displayed with the appropriate icon. Throughout the trial, this initial probability periodically changed (at random intervals between 30 and 90 seconds) to become more indicative of their actual identity (e.g., move more toward either 0% or 100% friendly) 70% of the time, and less certain 30% of the time. Thus, with time, identities became more and more certain; however, any one change was not completely predictive of the object's actual identity.

Each trial had a maximum time limit of ten minutes. Participants were assigned 60 resources at the beginning of each trial. Participants lost 1 resource for every correct identification, and 2 resources for every incorrect identification. Additionally, remaining resources corresponding to 16% of the unidentified contacts were lost each minute. Overall score per trial was computed as the number of correct identifications plus the resources remaining at the end of the trial, multiplied by a factor of 10. The intent of these manipulations was to reward participants' waiting until they were confident of an object's identity (so they wouldn't make an incorrect identification) but to not wait longer than necessary (or resources would be lost).

6.4.3 Preliminary Results

Overall score and time to complete trials have been examined. Figures 6.4 and 6.5 show score for each trial by display condition, and by picture condition, respectively. Figures 6.6 and 6.7 show time to complete trials by display condition, and picture condition, respectively.

Examination of these figures indicates that participants' performance, as measured by overall score, increased over the first three trials and tended to level out to trial four, while time to complete the trials tended to decrease over the four trials, indicating that participants tended to learn over the course of the experiment. There was little difference between picture conditions, indicating that changes in icons or icon categories did not affect performance. Additionally, there was little difference between

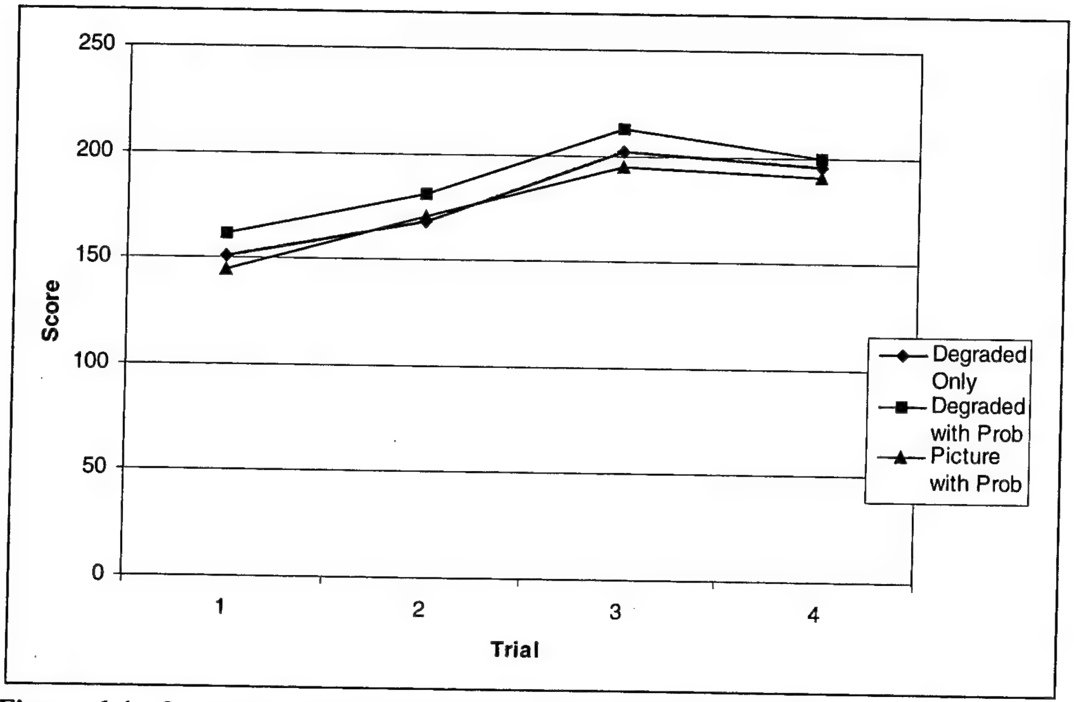


Figure 6.4. Overall score for each trial, by Display Condition.

display conditions, indicating that participants did not perform differently based on whether they saw just degraded images, degraded images combined with probabilities, or probabilities with non-degraded images.

Statistical results support these interpretations. Overall score and trial time were analyzed using a mixed model. Display type and picture were treated as between-subjects factors, and trial was treated as a within-subjects factors. There was a significant effect of trial on both score, *F*(3, 189) = 38:243, $p < .000$, and trial time, *F*(3, 189) = 9.94, $p < .000$, however, no other main effects or interactions approached significance.

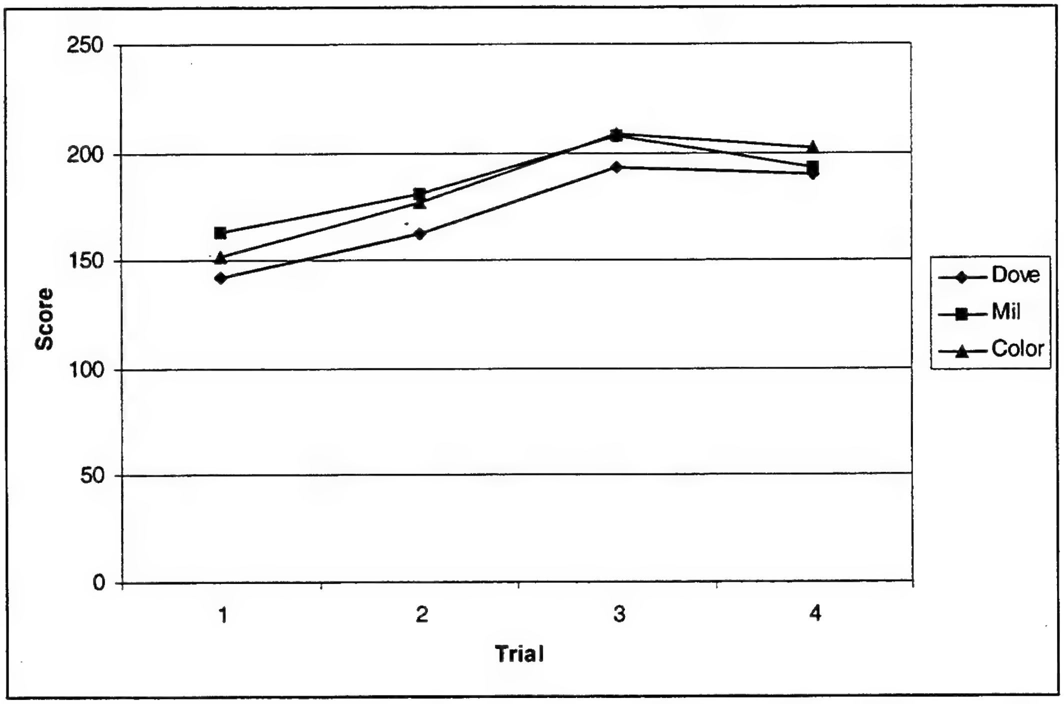


Figure 6.5. Overall score for each trial, by Picture Condition.

6.4.4 Discussion and Conclusions

Task performance results indicated that displays using degraded icons to convey uncertainty performed as well as degraded icons annotated with numeric probabilities, or in a situation where the only information regarding uncertainty was conveyed via numeric probabilities. That is, participants were able to perform equally well when their only information regarding object uncertainty was conveyed via the degraded images: the presence of numeric probabilities did not provide an advantage in this task. This result is significant, because it indicates both that people are able to understand uncertainty conveyed through such a manner, and thus that the use of distorted or degraded images may be a viable alternative to convey situational uncertainty. This may be particularly advantageous on displays, such as military tactical displays, where many objects may need to be displayed, and annotations of numeric probabilities may additionally clutter the display space.

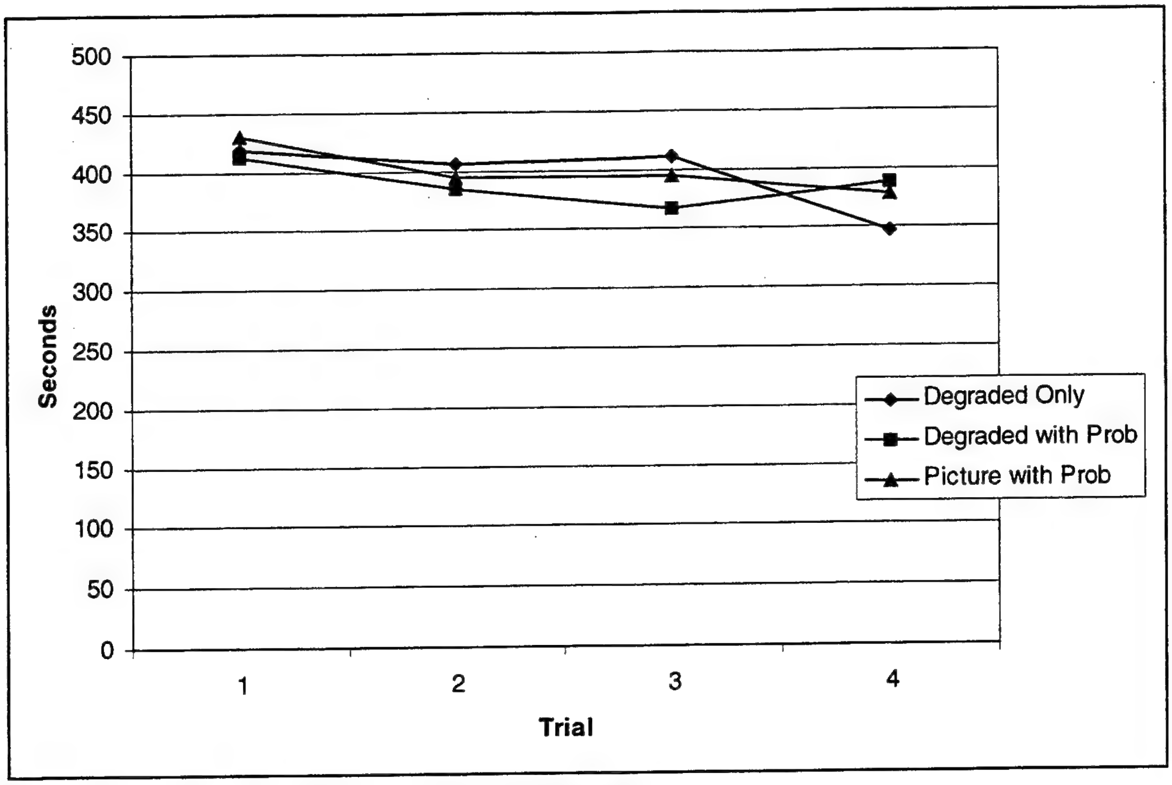


Figure 6.6. Time per trial, by Display Condition.

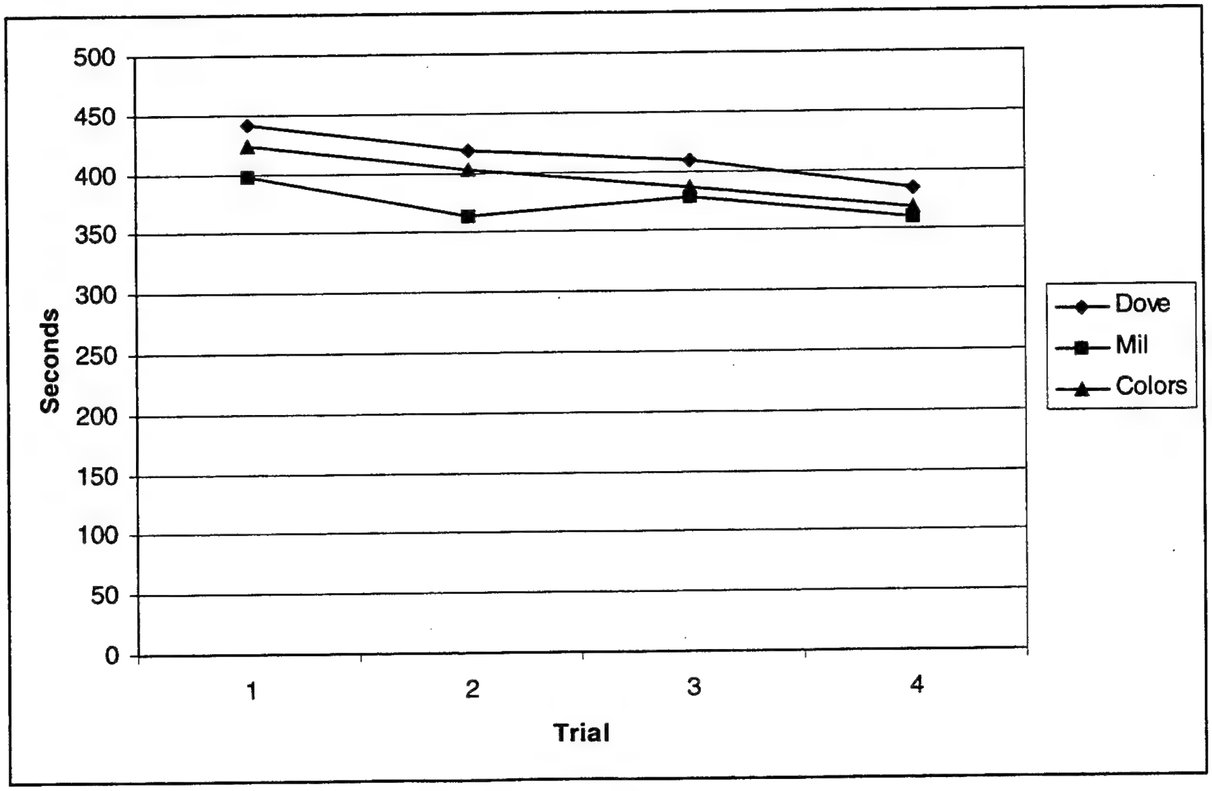


Figure 6.7. Time per trial, by Picture Condition.

CHAPTER 7

POTENTIAL FUTURE WORK

7.1 Further Experimentation

At a basic level, further experimentation in this area should focus on the continued investigation of human operators responses to test scenarios which stem from, and span, spaces in the theoretical framework described in Chapter 2. The studies described here focused on manipulations across one dimension - the attributed cause of the failure; future studies should investigate scenarios in which other dimensions are varied, either singly, or in combination to allow the sensitivity of human responses to combinations of factors to be tested.

7.2 Multi-crew Studies

Additionally, the framework described above can provide the basis for multiple participant studies, which investigate multi-decision maker, multi-team, or truly adversarial decision-making situations. In a multiple participant scenario, there are several possibilities. For example, two or more participants could collaborate against simulated hostile forces and information warfare attacks. In contrast to experiments with single participants, with multiple players it would be possible to investigate how people integrate information they obtain through their own systems with information they obtain "second-hand" from other people or teams, and how trust in and use of those information sources may be differentially affected by its source.

Utilization of multiple players in different locations with separate aiding systems will also allow investigation of situations in which information attacks may have affected information systems differentially (e.g., through disruption to information transmission or displays), and how participants act in the face of these discrepancies. Additionally, participants could compete against each other in an adversarial setting, with information warfare attacks introduced either automatically (i.e., experimentally controlled interventions through either a pre-set simulation or through an experimenter controlled test-controller station) or at the discretion of the participants.

7.3 Information Displays and Trust

The work described in Chapter 6 regarding information displays should be expanded and integrated with the empirical work on human trust in adversarial situations described in Chapters 4 and 5, and proposed above, in order to investigate how different displays of uncertain information may affect operator use of decision-aids, particularly under conditions where decision-aid outputs might be degraded or tampered with. In particular, the extent to which operators can detect information attacks, and learn to trust and use information aids, when uncertain estimates are provided graphically vs. through numeric probabilities, is of interest.

7.4 Increased Operational Realism

The studies presented here were conducted in the context of experimental micro-worlds with naïve participants. While a first step towards the examination of the new ideas regarding human trust and information displays was presented here, such studies need to incorporate increasing degrees of operational realism in order to be fully generalizable and applicable to real world situations. There are several steps that can be taken in future research to increase operational realism. For example, investigations could be conducted in the context of dynamic computer test-beds designed to incorporate and/or mimic operationally relevant displays and decision aids (e.g., data-fusion based aids) which may be impacted by adversarial information operations. Test scenarios could be developed based on actual operational situations. Finally, the participant pool should be expanded from naïve subjects to operators experienced with military environments and information operation situations.

7.5 Development of an AADM Laboratory Environment

Finally, to study these issues in a multi-player scenario, a fully functioning laboratory to study issues of group decision-making in adversarial situations such as those created through offensive and defensive information operations will need to be created. As noted in Section 7.2, experiments with multiple players will allow the investigation of how people integrate information across human and computerized sources, and how trust in and use of those information sources may be differentially affected by its source. To support such experimentation, which would include two adversarial teams composed of multiple operators (e.g., three-on-three experiments), an AADM laboratory would require a suite of interconnected experimental stations, in two separate locations (one for each team), a test-controller station to interject additional actions (e.g., a simulated hardware failure), and audio and video recording equipment for data capture. Such a laboratory could be developed in stages; for instance, through the development first of a test controller-experimental station pair to allow participants to "compete" against an intelligent (but experimentally controlled) adversary; to experimental workstation pairs to allow one-on-one experimentation; to two groups of multiple workstations.

CHAPTER 8

REFERENCES

Andre, A. D., & Cutler, H. A. (1998). Displaying Uncertainty in Advanced Navigation Systems. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (pp.31-35).

Banbury, S., Selcon, S., Endsley, M., Gorton, T., & Tatlock, K. (1998). Being Certain About Uncertainty: How the Representation of System Reliability Affects Pilot Decision Making. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (pp.36-39).

Brehmer, B. (1990). Strategies in Real-time, Dynamic Decision Making. In R. M. Hogarth (Ed.), *Insights in Decision Making,* Chicago: University of Chicago.

Brehmer, B. & Allard, R. (1991). Dynamic decision making: The effects of task complexity and feedback delay. In J. Rasmussen, B. Brehmer, & J. Leplat (Eds.), *Distributed Decision Making: Cognitive Models for Cooperative Work* (pp. 319 - 334). John Wiley and Sons.

Cannon-Bowers, J. A., Salas, E., & Pruitt, J. S. (1996). Establishing the boundaries of a paradigm for decision-making research. *Human Factors, 38*(2), 193 - 205.

Connolly, T. (1988). Hedge-clipping, tree-felling, and the management of ambiguity: The need for new images of decision-making. In L. R. Pondy, R. J. Boland. Jr., & H. Thomas (Eds.), *Managing Ambiguity and Change* (pp. 37 - 50). New York: Wiley and Sons.

Cooksey, R. W. (1996). *Judgment Analysis: Theory, Methods, ànd Applications.* San Diego: Academic Press.

Hammond, K. R., Stewart, T., Brehmer, B. & Steinmann, D. O. (1975). Social judgment theory. In M. F. Kaplan and S. Schwartz, Eds. *Human Judgment and Decision Processes.* New York: Academic Press.

Hogarth, R. M. (1981). Beyond Discrete Biases: Functional and Dysfunctional Aspects of Judgmental Heuristics. *Psychological Bulletin, 90*(2), 197 - 217.

Jian, J-Y, Bisantz, A. M., & Drury, C. G. (2000). Foundations for an Empirically Determined Scale of Trust in Automated Systems, *International Journal of Cognitive Ergonomics, 4*(1), 2000. 53-72.

Kantowitz, B. H., Hanowski, R. J., & Kantowitz, S. C. (1997). Driver Acceptance of Unreliable Traffic Information in Familiar and Unfamiliar Settings. *Human Factors, 39*(2), 164 – 176.

Kirschenbaum, S. S., & Arruda, J. E. (1994). Effects of Graphic and Verbal Probability Information on Command Decision Making. *Human Factors, 36*(3), 406 – 418.

Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making. In G. A. Klein, J. Orasanu, R. Calderwood, C. E. Zsambok (Eds.), *Decision Making in Action: Models and Methods* (pp. 138 - 147). Norwood, NJ: Ablex.

Klein, G. A. & Calderwood, R. (1991). *Decision Models: Some Lessons From the Field. IEEE Transactions on Systems, Man, and Cybernetics, 21*(5), 1018 – 1026.

Larzelere, R. E., & Huston, T. L. (1980). The Dyadic Trust Scale: Toward Understanding Interpersonal Trust in Close Relationships. *Journal of Marriage and the Family*, 595-604.

Lee, J. D. & Moray, N. (1994). Trust, Self-confidence, and Operators' Adaptation to Automation. *International Journal of Human-Computer Studies, 40*, 153-184.

Lee, J. D., & Moray, N. (1992). Trust, Control Strategies and Allocation of Function in Human-machine Systems. *Ergonomics, 35*(10), 1243-1270.

Lerch, F. J., & Prietula, M. J. (1989). How do we trust machine advice? In G. Salvendy, and M.J. Smith, (Eds.) *Designing and Using Human-computer Interface and Knowledge Based Systems*. North-Holland: Elsevier Science Publishers.

Lind, A. T., Dershowitz, A., Chandra, D., & Bussolari, S. R. (1995). The effect of data link-provided graphical weather images on pilot decision making. In *IFAC Proceedings* – 1995.

Llinas, J., Bisantz, A. M., Drury, C. G., Seong, Y., & Jian, J-Y. (1998). *Studies and Analyses of Aided Adversarial Decision-making. Phase 2: Research on Human Trust in Automation.* Center for Multi-Source Information Fusion Technical Report. State University of New York at Buffalo, Amherst, NY.

Llinas, J., Drury, C., Bialas, W., Chen, A. C. (in press). Studies and Analyses of Vulnerabilities in Aided Adversarial Decision Making.*(AFRL/HE-WP-TR-1998-0099)* Wright-Patterson AFB, OH: Air Force Research Laboratory, Human Effectiveness Directorate, Crew System Interface Division.

Muir, B. M., & Moray, N. (1996). Trust in Automation: Part II. Experimental Studies of Trust and Human Intervention in a Process Control Simulation. *Ergonomics, 39*(3), 429-460.

National Research Council (1997). Automation. In C. D. Wickens, A. S. Mavor, & J. P. McGee, (Eds.). *Flight to the Future: Human Factors in Air Traffic Control.* (pp.271-289). National Academy Press, Washington D.C.: Author.

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance Consequences of Automation-induced "Complacency." *The International Journal of Aviation Psychology, 3*(1), 1-23.

Rasmussen, J., Pejtersen, A. & Goodstein, L. P. (1994). *Cognitive Systems Engineering.* New York: Wiley and Sons.

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in Close Relationships. *Journal of Personality and Social Psychology, 49*(1), 95-112.

Savage, L. (1954). *The Foundation of Statistics*. New York: Wiley and Sons.

Sheridan, T. B. (1988). Trustworthiness of Command and Control Systems. *IFAC Man-Machine Systems*, 427-431.

Simon, H. A. (1960). *The New Science of Management Decision*. New York: Harper and Brothers.

Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Automation-induced "Complacency": Development of the Complacency-Potential Rating Scale. *The International Journal of Aviation Psychology, 3*(2), 111-122.

Von Neumann, J. & Morgenstern, O. (1944). *Theory of Games and Economic Behavior.* Princeton: Princeton University Press.

Walts, E. & Llinas, J. (1990). *Multi-sensor Data Fusion.* Norwood, MA: Avtech House.

Zhang, L., Helander, M. G., & Drury, C. G. (1996). *Identifying Factors of Comfort and Discomfort in Sitting. Human Factors, 38*(3), 377-389.

Zsambok, C. E. & Klein, G. (1997). *Naturalistic Decision Making.* Mahwah, NJ: Lawrence Erlbaum.

APPENDIX A

The following paper was presented at Fusion '99, in Sunnyvale, California, July, 1999.

Human Performance and Data Fusion Based Decision Aids

Ann M. Bisantz , Richard Finger, Younho Seong, and James Llinas
Center for Multi-source Information Fusion
Department of Industrial Engineering
State University of New York at Buffalo
Amherst, NY 14260
bisantz@eng.buffalo.edu

Abstract – Decision-aids based on data fusion technologies may be applied to support decision-making in a variety of environments, ranging from military command and control situations to intelligent transportation applications. In any situation, the ultimate performance of human decision-maker/decision-aid system depends not only on the quality of the aid, but on the human decision-maker's utilization of the information provided by the aid. This utilization can be affected by many factors, including the degree of trust the decision-maker has in the aid, and the form in which information is presented to the decision-maker. This paper describes a framework for investigating trust in data-fusion based decision aids, and results from a pilot experiment in which distorted and blended graphical forms were used to represent uncertain information.

Key Words: decision-aids, trust, information displays.

1. Introduction

1.1 Data Fusion Based Decision-Aids

Decision-aids based on data fusion technologies may be applied to support decision-making in complex, dynamic environments such as military command and control, non-destructive testing and maintenance, and intelligent transportation. These aids provide operators with situational estimates which can aid in the decision-making process. For instance, in a military environment, data fusion based decision-aids may provide commanders with estimates of an entity's identity or threat potential. Regardless of environment, such aids provide decision-makers with information that has an associated level of confidence or uncertainty, through the application of automated algorithms and processes. The ultimate performance of such systems, consisting of both the human decision-maker and the automated decision-aid, depends on the human decision-makers' utilization of the information provided by the aid. Such utilization can be impacted by many factors, including the level of risk, time pressure, nature of the information display, and level of trust the decision-maker has in the automated aid.

This paper describes a research approach addressing the latter two factors in the context of a military environment. In a military context, data fusion has been identified as a means to perform assessments of identities, situations, and threat potential based on information derived from multiple electronic and intelligence sources. In these situations, the inherent risks, time pressure and large volume of data have led to the need for computerized aids performing automated data fusion (Walts and Llinas, 1990).

The process of data fusion in a military context includes multiple levels, each of which provides information at a different level of abstraction. For instance, different levels would address the detection and identification of potential targets, the association of targets into organized groups with certain behaviors, and the estimation of the threat potential of those groups. Thus, the results of data fusion processing can provide input to the situation assessment activities of battlefield commanders (Llinas, Drury, Bialas, and Chen, in press). Ultimately, information resulting from the data fusion process is presented to the human decision-maker through a computer interface.

1.2. Decision Aiding in an Adversarial Environment

Aided-adversarial decision-making (AADM) refers to military command and control decision making in environments in which computerized aids are available, and in which there is a potential for adversarial forces to tamper with and disrupt such aids. Hostile forces may attempt to compromise tactical decision-making through offensive activities conducted to attack or interfere with an adversary's information systems. Information warfare can impact an adversary's operations through information disruption, denial, and distortion (Llinas, Drury, Bialas, and Chen, in press). For instance, disrupting or denying access to sources of information may make it difficult for decision-makers to assess situations and take appropriate actions. Distorted information, through manipulation and addition of incorrect information, may fool adversaries into taking actions desirable from a friendly perspective.

1.3 Human Trust in Automated Aids

Given the potential for information operations to disrupt and corrupt information provided by data-fusion based aids, it is necessary to understand the extent to which decision-makers rely on or use these aids, and factors affecting that reliance. A possible source of information regarding these issues is research that has been performed in the area of human trust in automated systems (e.g., Lee and Moray, 1992; Muir and Moray, 1996; Parasuraman, Molloy, and Singh, 1993; Sheridan, 1988). Researchers have suggested that trust can affect how much people accept and rely on increasingly automated systems (Sheridan, 1988).

Generally, research from both social science and engineering perspectives agree that trust is a multi-dimensional, dynamic concept capturing many different notions. For example, Rempel et al. (1985) concluded that trust would progress in three stages over time from predictability, to dependability to faith. Muir and Moray (1996) extended these three factors, and developed an additive trust model that contained six components: predictability, dependability, faith, competence, responsibility, and reliability. Sheridan (1988) also suggested possible factors in trust, including reliability, robustness, familiarity, understandability, explication of intention, usefulness, and dependence.

Empirical results have shown that people's strategies with respect to the utilization of an automated system may be affected by their trust in that system. For example, Muir and Moray (1996) and Lee and Moray (1994) studied issues of human trust in simulated, semi-automated pasteurization plants. These studies showed, among other results, that operators' decisions to utilize either automated or manual control depended on their trust in the automation and their self confidence in their own abilities to control the system. Additionally, results showed that trust depended on current and prior levels of system performance, the presence of faults, and prior levels of trust. For example, trust declined, but then began to recover, after faults were introduced (Lee and Moray, 1992). Lerch and Prietula (1989) found a similar pattern in participants' confidence in a system for giving financial management advice: confidence declined after poor advice was given, then recovered, but not to the initial level of confidence.

In the context of AADM, there exists the potential for several circumstances in which trust in data-fusion based decision aids could be affected. For instance, information warfare techniques could be used by an adversary to distort the information provided by decision aiding systems, disrupting (appropriately) commanders' trust in, and utilization of, such systems. Alternatively, an adversary might act deceptively, fooling a commander into trusting and acting based on information in a way favorable to the adversary. Finally, an adversary might disrupt a commander's trust in an aid that is providing good ("trustworthy") information. For these reasons, it is necessary to investigate human trust in AADM situations, in order to better understand how data-fusion based decision aids will impact the decision-making process under different circumstances.

2.0 Investigations of Decision Aiding in Adversarial Environments

2.1 Theoretical Framework

To structure the investigation of aspects of human trust in data fusion-based decision aids, a multi-dimensional framework was developed (Llinas, Bisantz, Drury, Seong, and Jian, 1998). The framework integrates and systematically varies a set of dimensions which may affect trust in decision aids. The following dimensions are included in the framework:

1. Locus of Attack. One potential factor is the location at which the potential for corruption exists. Two potential dimensions can contribute to this factor: the component dimension, and the surface-depth dimension.
a) Component Dimension. Information could be corrupted at a variety of components, or levels, in the AADM environment. Information could be corrupted at the level of the tactical situation (by interfering with sensors), within the information processing and data fusion algorithms that comprise the decision aids, or at the level of the human-computer interface.
b) Surface-Depth Dimension. A second related dimension along which investigations of performance in AADM systems can vary is a surface-depth dimension. The surface level corresponds to the information available about the environment (as formalized in Brunswik's Lens Model; Cooksey, 1996; Hammond, Stewart, Brehmer, and Steinman, 1975), whereas the depth level corresponds to the actual state of the environment. In an AADM environment, surface level features would be the observable outputs from sensors, or data fusion processes. Depth level features would be the actual operations of the sensors or algorithms themselves.
2. Malfunction Level. Information aids for AADM can fail or be corrupted in qualitatively different ways, either failing completely, or being partially degraded, resulting in two malfunction levels:

- Element failure. System components can fail completely resulting in a loss of data.
- Element degradation. The quality of information provided by the system component can be degraded, resulting in partial information loss, or increased ambiguity and uncertainty.

3. Causes of Failure or Corruption. Information can be corrupted through different causes or intentions, ranging from naturally occurring system failures (e.g., hardware malfunctions), to deliberate attacks on the information systems, to deliberate attacks which are disguised by the adversary.
4. Time Patterns of Failure. A final dimension reflects the dynamic or time-dependent characteristics of the degradation. Failures, sabotage, and subterfuge can occur not only as failures or degradations at a particular point in time, but also in a continuing fashion. Additionally, failures can occur with patterns that are either predictable or unpredictable.

2.2 Framework-based Experiments

This framework is being used to develop experiments in the area of human trust in data fusion-based decision aids. At present, experiments are planned to investigate changes in trust in, and reliance on, a data-fusion aid when the situation is framed as either one in which the aid may be unreliable due to hardware failures, or one in which the aid may be subject to deliberate tampering by an adversary. Participants will perform a simulated military command and control task in which they will identify unknown aircraft moving on a radar screen.

During the task, participants will be able to access both non-aid information (e.g., altitude, radar emission, and speed information) about unknown aircraft, as well as an identity estimate from a simulate data-fusion aid. The identity estimate will be in the form of a probabilistic range (e.g., that an aircraft is friendly or hostile). Participants will request access to either type of information, and will be limited in the number of requests, forcing a tradeoff between information sources.

Participants will perform the experiment over six scenarios, during which time the speed and altitude of the aircraft will vary within pre-defined, overlapping ranges. Prior to the experiment, participants will be given conditional probability information about the chance that an aircraft is hostile, given that it is flying at a particular speed altitude, or has a particular radar signature.

After several three normal scenarios, a fault (either a constant shift in the probabilistic range, or a gradually increasing range) will be introduced into range provided by the data-fusion aid. Participant's reliance on either form of information (either the decision aid, or the other available information) will be measured before and after the insertion of the error to assess the potential loss of trust in the aid subsequent to the error.

3.0 Investigations of Data Presentation

As noted above, one factor which may influence the utility of data fusion based decision aids, and the influence of these aids on the decision making process, is the form in which the uncertain information determined by these aids is presented to decision-makers. Uncertain or probabilistic Information can be shown in a variety of formats ranging from simply text to graphical representations to text/graphical hybrids. Past research has focused on representing position, direction and identity uncertainty in a format that reveals the true probabilistic nature behind the data (Andre and Cutler, 1998; Banbury, Selcon, Endsley, Gorton, and Tatlock, 1998; Kirschenbaum and Arruda, 1994).

Position uncertainty deals with how to represent the possible places an object may inhabit. Environments in which this type of uncertainty plays an important role include commercial aviation and military sonar/radar. Andre and Cutler(1998) investigated this form of uncertainty with the use of a task in which a pilot would have to play "Chicken" with a circular object, they called a meteor. The pilot's goal was to come as close as possible to the meteor without collision. To represent the position uncertainty a circular ring surrounded the meteor. The ring varied in size dependent upon uncertainty level. Collision frequency was found to be far less when the ring was displayed: without the ring, participants appeared to dismiss the fact that uncertainty was present in the system. Kirschenbaum and Arruda (1994) conducted a similar experiment which investigated the effect of different displays of position uncertainty on a decision-making task as to when and where to fire at a target. Participants were shown either a graphical representation of position uncertainty in the form of an ellipse around the target or a verbal indicator that ranged from poor to fair to good. The elliptical aid was found to be superior to the verbal in cases of moderate to high difficulty scenarios. Overall it appears that the use of a visual position uncertainty aid helped the performance of the user.

Aids which present heading uncertainty attempt to display all the possible future directions an object may move. Andre and Cutler (1998) tested three different types of heading uncertainty aids in a simulated anti-aircraft task: a textual description and two graphical representations that utilized either arcs or rings. The three aids improved user performance when compared with a no aid condition. The arc-based aid, which represented the uncertainty in direction by utilizing an arc that covered the entire angle of possible movement heading, provided a slight advantage over the other two aids.

Finally, identity aids strive to give the user an idea of how accurate the identification of an object is. Currently most aids display this information in the form of probabilities. Banbury et al. (1998) investigated how the context in which information is displayed affects a decision-making task. Participants were asked to make a shoot/no-shoot decision based on a probabilistic estimate of an aircraft's identity, presented as a numeric percentage. Results showed an impact of estimate uncertainty - participants were found to have a reluctance to shoot when uncertainty was greater than 9%. Additionally, presenting a secondary target identification (e.g., not just the chance that is a hostile fighter, but also the chance that it is a friendly aircraft) also impacted decisions to shoot. Participants were more hesitant when a secondary, friendly, target identification estimate was given.

Another way in which the graphical form of information presentation could be used to represent uncertainty is through the use of degraded or distorted images. Lind, Dershowitz, Chandra, and Bussolari (1995) provide evidence that the form of displayed information may affect the use of uncertain data. In a study to investigate the extent to which the graphic depiction of weather systems could be degraded (due to technical limitations) and still be acceptable to general aviation pilots, Lind et al. found that pilots' estimates of weather hazards increased as the graphical distortion increased. In this case, the distortion

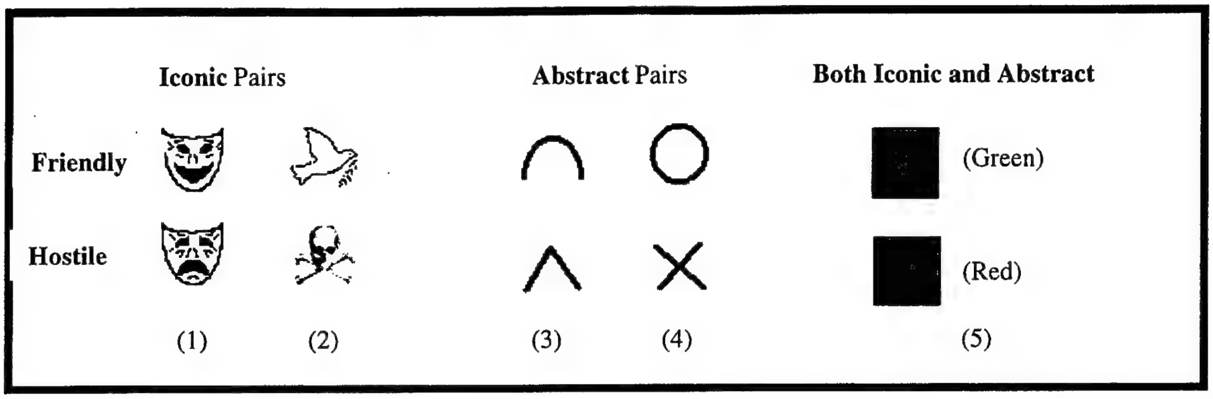


Figure 1. Five pairs of icons representing object identities as either hostile or friendly.

took the form of larger polygon/ellipse shaped depictions of weather patterns, in contrast to the non-distorted continuous, fine-grained representation. This increase in perceived risk might indicate a decrease in subjects' confidence of their understanding of the current specific weather patterns.

Thus, there is some indication that iconic representations based on degraded or distorted images may be used to convey the uncertainty associated with a decision aid estimate. In the following pilot study, we investigated properties of distorted and blended icon sets intended to convey uncertain information about an object's identity as either potentially hostile or friendly. Future experiments will investigate the impact of a subset of these icons, selected based on the pilot study results, on a decision-making task.

3.1 Pilot Study Method

3.1.1 Participants

Twenty participants, all undergraduate students, were paid $6.00 per hour for their participation in the pilot study.

3.1.2. Experimental Design

Five sets of pictures were chosen to represent the identity of an object as either hostile or friendly. These picture sets were classified as either abstract (without an obvious associated meaning), iconic (with an associated meaning), or both. Picture pairs were chosen in order to allow for the entire spectrum from friendly to hostile to be represented. Figure 1 shows the pictures used in the experiment.

In order to represent the probabilistic nature of the information graphically, a series of thirteen icons were created to represent a range of probabilities (i.e., from p(Hostile) – 0.0 to p(Hostile) = 1.0). The iconic and abstract picture pairs were distorted and blended using a pixelizing function found in Adobe PhotoShop 4.0. For example, the 50% friendly/50% hostile picture blended both of the pictures in a pair together. For the colored icons, the series of icons was created by coloring each pixel in the icon as either green or red based upon the probability desired. To illustrate how the pixelizing function works, the series of the distorted and blended pictures for picture pair (1) are shown in Figure 2.
Each participant performed a series of tasks involving all five sets of icons. Ten participants performed the tasks under a "friendly" framing condition, and ten participants performed the tasks under a "hostile" framing condition. In the friendly framing condition, participants were given task instructions which described the icons as more or less friendly. In the hostile framing condition, icons were described as more or less hostile.

3.1.3 Procedure

The three experimental tasks were designed to measure whether the icons could be correctly sorted and assigned a probability rating according to the expected probabilities that the icons represented. Participants performed each of the tasks five times: once for each icon pair (see Figure 1).

In the first task, a timed sorting task, participants were asked to sort cards into piles according to the icon printed on the card. Participants were asked to create piles containing the same icon. There were five instances each of the 13 possible icons in a set, for a total of 65 cards. The time to sort the cards, and sorting errors, were collected.

In the second task, participants were asked to order the set of thirteen pictures from most to least friendly (or hostile), depending on the framing condition. They were not told which icons corresponded to the hostile or friendly ends of the scale (e.g., they were not told that a circle represented a most friendly, and an "x", least friendly). Participants performed this task using a Visual Basic computer program, through which they could drag and drop the icons into the desired order. The ordering of the icons was recorded automatically by the computer.

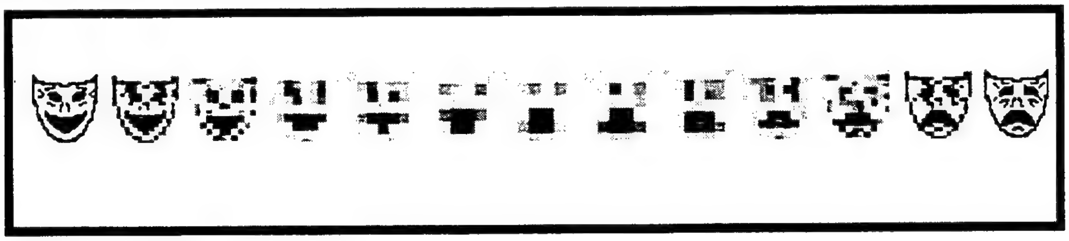


Figure 2. Series of 13 icons representing a range of probabilities that an object is hostile or friendly: from a probability of 100% friendly to 100% hostile.

For the third task, participants were asked to rate each icon on continuous scale, with end points of least and most friendly (or hostile). Participants marked their rating along a line connecting the endpoints; this distance was later measured and scaled based on the length of the line, and used to identify their rating.

3.2 Pilot Study Results

3.2.1 Card Sorting

The times to sort cards based on the icon printed on the card did not differ significantly across picture pairs. Thus, the relative difficulty of identifying and sorting the thirteen icons did not appear to differ across sets.

3.2.2. Ordering

The order of the thirteen icons in each icon pair set was determined for each participant, for the hostile and friendly framing conditions, resulting in ten orders per icon pair for each framing condition. These orders were used to compute an average ranking for each icons for the five pairs, for both framing conditions. Ordering these average rankings resulted in an average order for each set, for both framing conditions (a total of 10 average orders). These average orders were correlated with the expected order (based on the way the icons were created), and a Spearman correlation coefficient was computed. These coefficients are shown in Table 1. All correlations were significant at the .01 level of significance, indicating that overall, participants were able to correctly order the sets of icons according to the intended levels of uncertainty.

Table 1. Spearman Correlation Coefficients comparing average rank orders to expected order for 5 Icon Pairs.

	Framing	
Icon Pair	Friendly	Hostile
Mask(1)	1.000	0.929
Dove(2)	1.000	0.984
Inverted V-U (3)	0.934	0.984
Circle-X(4)	1.000	0.951
Color(5)	1.000	0.890

Individual participant data was also examined: Spearman correlation coefficients were computed comparing each participant's order to the expected order, for both framing conditions. These correlations are indicated in Tables 2 and 3, corresponding to the Friendly and Hostile framing conditions, respectively. Correlations in bold are *insignificant* at the .05 level of significance. Inspection of Tables 2 and 3 shows that on a participant-by-participant basis, ordering was more consistent and correct in the friendly framing condition than the hostile framing condition. Note that negative correlations simply indicate that the participant reversed the hostile and friendly ends of the scale (they were not told which icons corresponded to which endpoints before the experiment). It is interesting to note that even for the two "abstract" icons, reversals happened at a rate less than chance, indicating that perhaps there was some meaning intrinsic to the abstract icons.

Table 2. Individual Correlation Coefficients for each participant (Friendly framing condition; bold correlations are *insignificant*).

P's	Mask	Dove	V-U	Circle	Color
	(1)	(2)	(3)	(4)	(5)
2	0.995	1.000	-1.000	1.000	0.995
4	0.989	0.995	1.000	1.000	1.000
6	1.000	1.000	0.995	1.000	1.000
8	1.000	1.000	-1.000	1,000	0.995
10	1.000	1.000	1.000	1.000	1.000
12	1.000	-1.000	-1.000	-1.000	-1.000
14	0.984	1.000	1.000	1.000	0.995
16	0.962	1.000	1.000	1.000	1.000
18	0.995	0.995	1.000	1.000	1.000
20	0.978	1.000	**-0.440**	**0.374**	1.000

Table 3. Individual Correlation Coefficients for each participant (Hostile framing condition; bold correlations are *insignificant*).

P'	Mask	Dove	V-U	Circle	Color
s	(1)	(2)	(3)	(4)	(5)
1	**0.126**	**0.115**	**0.115**	0.115	**0.115**
3	1.000	1.000	1.000	1.000	0.995
5	0.566	**-0.038**	0.544	**-0.297**	0.665
7	**0.412**	**0.093**	**0.115**	**0.148**	**0.088**
9	**0.005**	0.714	**0.099**	**0.044**	**0.181**
11	0.978	1.000	1.000	1.000	**0.434**
13	**0.148**	1.000	1.000	0.995	**0.456**
15	0.978	1.000	0.978	0.995	0.989
17	0.995	0.995	1.000	1.000	1.000
19	**-0.165**	0.516	**0.280**	**0.440**	0.835

3.2.3 Rating

From the data collected on individual picture ratings an average rating was calculated for each picture within a picture pair category. These averages provided a range of estimates of the friendliness or hostility of each picture pair (Tables 4 and 5).

Table 4. Rating Spread for 5 icon pairs (Friendly Framing)

	Mask (1)	Dove (2)	V-U (3)	Circle (4)	Color (5)
High Rating	88.67	97.93	96.64	97.73	98.59
Low Rating	4.22	4.06	3.67	8.52	11.33

Note: Ratings for Dove, V_U, and Circle were corrected to account for obvious and consistent reversals between hostile and friendly endpoints.

Table 5. Rating Spread for 5 icon pairs (Hostile Framing)

	Mask (1)	Dove (2)	V-U (3)	Circle (4)	Color (5)
High Rating	66.56	72.27	62.11	74.06	63.69
Low Rating	24.22	17.97	38.20	21.48	14.30

3.2 Future Experiments

Future experiments will test the effect a subset of these icon pairs on decision-making in a dynamic identification task. Participants will be asked to identify objects as either friendly or not friendly, given a graphical icon of the object which depicts a decision-aid's probabilistic estimate of the object's identity. This icon will be based on either end-point icons with associated numeric probabilities, the full range of 13 icons, or the full range of 13 icons with associated numeric probabilities. Over time, estimates will tend (with some randomness) to become more certain; however, participants will be penalized for identification delays. The experiments will investigate the impact of information presentation on the point at which participants choose to identify objects. If graphical depictions (i.e., distorted icons) convey more information about the probabilistic nature of the identity estimate than numeric probabilities, then participants seeing the graphical depictions should choose to wait to make an identification until they are more certain.

4.0 Conclusions

Data fusion-based decision-aids can be implemented to provide support in a variety of situations. In order for those aids to provide effective support, the must provide information in a format that conveys important aspects of that information (e.g., its uncertain nature) and be trusted by the decision-maker. A framework for investigating trust in decision –aids, in adversarial decision-making situations, along with on-going experiments based on that framework, was discussed. Additionally, results from a pilot study investigating the utility of degraded and distorted images to convey levels of uncertainty were presented. Preliminary results indicated that sets of distorted icons could be appropriately ordered, and span a range of descriptive level, under particular framing conditions. Future experiments to investigate the effect of these representations on decision-making were described.

5.0 References

[1]. Andre, A. D., and Cutler, H. A. (1998). Displaying Uncertainty in Advanced Navigation Systems. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (pp.31-35).

[2] Banbury, S., Selcon, S., Endsley, M., Gorton, T., and Tatlock, K. (1998). Being Certain About Uncertainty: How the Representation of System Reliability Affects Pilot Decision Making. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (pp.36-39).

[3] Kirschenbaum, S. S., and Arruda, J. E. (1994). Effects of Graphic and Verbal Probability Information on Command Decision Making. *Human Factors*, 36(3), 406 – 418.

[4] Lee, J. D., Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35(10), 1243-1270.

[5]Lee, J. D. and Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies*, 40, 153-184.

[6] Lerch, F. J., and Prietula, M. J. (1989) How do we trust machine advice? In Salvendy, G. and Smith, M. J. (Eds.) *Designing and using human-computer interface and knowledge based systems*. Elsevier Science Publishers, North-Holland.

[7] Lind, A. T., Dershowitz, A., Chandra, D., & Bussolari, S. R. (1995). The effect of data link-provided graphical weather images on pilot decision making. *In IFAC proceedings – 1995.*

[8] Llinas, J. Bisantz, A. M., Drury, C. G., Seong, Y., Jian, J-Y. (1998). *Studies and analyses of aided adversarial decision-making. Phase 2: research on human trust in automation.* Center for Multi-Source Information Fusion Technical Report. State University of New York at Buffalo, Amherst, NY.

[9] Llinas, J., Drury, C., Bialas, W., Chen, A. C. (in press). *Studies and Analyses of Vulnerabilities in Aided Adversarial Decision Making.*(AFRL/HE-WP-TR-1998-0099) Wright-Patterson AFB, OH: Air Force Research Laboratory, Human Effectiveness Directorate, Crew System Interface Division.

[10] Muir, B. M., Moray, N. (1996). Trust in automation: Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, 39(3), 429-460.

[11]Parasuraman, R., Molloy, R., Singh, I. L. (1993). Performance consequences of automation-induced "complacency". *The International Journal of Aviation Psychology*, 3(1), 1-23.

[12] Rempel, J. K., Holmes, J. G., and Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology*, 49(1), 95-112.

[13] Sheridan, T. B. (1988). Trustworthiness of command and control systems. In *IFAC Man-Machine Systems*, 427-431, Oulu, Finland.

[14] Walts, E. and Llinas, J. *Multi-sensor Data Fusion.* Norwood, MA: Avtech House.

Appendix B Scenario Files

Scenario 1

		Identity		Position						Radar		Conf. Interval	
ID	Type	curr	true	x	y	Speed	Heading	Altitude	Radar	on/off	sel	min	max
JC443	3	7	1	-193.25	-217.44	0.242	55	31293	APG-65	0	0	0.721	0.885
JC190	3	7	1	-296.37	-39.331	0.186	63	29098	APG-65	0	0	0.73	0.909
JC636	3	7	1	-69.884	-54.697	0.2	58	29242	APG-65	1	0	0.768	0.913
JC908	3	7	1	-276.07	124.63	0.253	286	20206	APQ-159	1	0	0.919	1
JC895	3	7	1	257.134	-102.41	0.113	87	13394	APG-73	1	0	0.906	1
JC962	3	7	1	92.3887	-72.275	0.219	263	33267	APG-73	1	0	0.509	0.649
JC113	3	7	1	-28.794	86.3598	0.446	20	10037	APQ-159	1	0	0.938	1
JC466	3	7	1	162.539	-29.473	0.187	74	29910	APG-65	1	0	0.767	0.895
JC876	3	7	1	77.2271	182.539	0.046	294	39648	APQ-159	0	0	0.934	1
JC225	3	7	1	-26.945	-89.457	0.264	357	30080	APQ-159	0	0	0.836	0.952
JC320	3	7	1	9.58586	143.597	0.29	275	33011	APG-65	1	0	0.765	0.906
JC141	3	7	1	204.984	-222.46	0.291	6	40160	APG-65	0	0	0.915	1
JC129	3	7	1	-27.805	-241.35	0.309	83	22100	APQ-159	1	0	0.929	1
JC248	3	7	1	233.11	-40.368	0.303	50	25169	APQ-159	1	0	0.792	0.988
JC297	3	7	1	-166.71	-246.17	0.142	193	2528	APQ-159	0	0	0.907	1
JC115	3	7	1	-269.18	5.74511	0.204	252	34521	APQ-159	0	0	0.789	0.959
JC356	3	7	1	153.932	-158.12	0.165	132	30435	APQ-73	1	0	0.435	0.585
JC048	3	7	1	64.2079	39.9411	0.046	292	33941	APQ-73	0	0	0.266	0.396
JC598	3	7	1	-133.81	4.47859	0.222	212	16466	APG-65	1	0	0.718	0.866
JC421	3	7	1	250.267	-217.6	0.062	240	28025	APR-25	1	0	0.919	1
JC434	3	7	1	10.831	152.264	0.294	225	32707	APR-25	1	0	0.801	0.966
JC420	3	7	1	-14.42	42.9319	0.142	38	36968	APQ-159	1	0	0.796	0.939
JC918	3	7	1	-136.21	148.724	0.247	30	34780	APG-73	0	0	0.494	0.655
JC519	3	7	1	267.04	25.9026	0.043	69	28680	APQ-159	1	0	0.92	1
JC483	3	7	1	143.092	197.737	0.197	253	25254	APG-73	1	0	0.495	0.636
JC373	3	7	1	-50.841	-246.63	0.244	281	40329	APQ-159	1	0	0.903	1
JC977	3	7	1	205.057	-90.876	0.186	98	32513	APQ-159	0	0	0.824	0.94
JC825	3	7	1	126.704	-74.015	0.316	252	35963	APG-65	1	0	0.749	0.889
JC991	3	7	1	94.8241	198.958	0.043	255	16466	APQ-73	1	0	0.235	0.415
JC330	3	7	1	-233.37	-140.77	0.142	6	37857	APR-25	1	0	0.874	1
JC956	3	7	1	-204.38	46.5484	0.264	297	36867	APG-65	1	0	0.757	0.894
JC148	3	7	1	-293.37	-41.803	0.32	65	26537	APG-73	1	0	0.947	1
T744	9	7	4	252.412	-28.112	0.168	15	36769	APG-73	1	0	0.35	0.474
T414	9	7	4	-178.13	18.2648	0.31	14	41555	APQ-159	0	0	0.92	1
T746	9	7	4	-179.13	-200.12	0.2	83	26125	APQ-73	1	0	0.106	0.298
T740	9	7	4	232.865	-20.112	0.147	222	33300	APG-73	1	0	0.369	0.49
T633	9	7	4	245.088	-51.265	0.34	146	37000	APR-25	1	0	0.907	1
T759	9	7	4	218.205	-177.97	0.18	325	33192	APG-145	1	0	0.553	0.668
T410	9	7	4	-199.13	21.2648	0.33	156	43212	APR-25	0	0	0.948	1
T415	9	7	4	-188.13	21.2648	0.33	180	46063	APR-25	0	0	0.939	1
T430	9	7	4	-187.95	28.5981	0.32	29	40000	APQ-159	1	0	0.936	1
T417	9	7	4	-179.13	24.5981	0.322	17	43718	APG-65	0	0	0.922	1
T431	9	7	4	-195.84	30.5981	0.19	84	38000	APQ-73	1	0	0.431	0.568
T422	9	7	4	-173.13	-51.265	0.2	52	25922	APG-145	1	0	0.52	0.689
T639	9	7	4	-168.13	-170.47	0.2	176	36544	APQ-73	1	0	0.429	0.565
T429	9	7	4	-180.07	26.5981	0.22	78	38000	APG-73	1	0	0.321	0.51
T526	9	7	4	-189.13	-93.843	0.385	135	42486	APR-25	0	0	0.9	1

Scenario 2

		Identity		Position						Radar		Conf. Interval	
ID	Type	curr	true	x	y	Speed	Heading	Altitude	Radar	on/off	sel	min	max
JC505	3	7	1	287.53	-73.13	0.268	60	32140	APQ-159	0	0	0.83	0.946
JC612	3	7	1	219.12	133.77	0.246	242	24469	APQ-159	1	0	0.91	1
JC613	3	7	1	147.36	-228.82	0.253	85	25196	APG-65	1	0	0.765	0.912
JC203	3	7	1	212.29	-198.7	0.185	59	39553	APG-73	1	0	0.5	0.647
JC222	3	7	1	-3.1404	-57.764	0.51	341	27204	APG-65	1	0	0.415	0.575
JC939	3	7	1	37.767	238.11	0.312	176	37959	APQ-159	1	0	0.832	0.958
JC629	3	7	1	269.57	-211.47	0.317	124	39655	APG-73	1	0	0.53	0.632
JC237	3	7	1	-251.31	72.504	0.332	296	32302	APG-73	1	0	0.508	0.669
JC885	3	7	1	243.55	154.97	0.488	41	16311	APQ-159	0	0	0.908	1
JC278	3	7	1	-167.35	170.12	0.241	309	28216	APG-73	1	0	0.527	0.651
JC615	3	7	1	-266.62	83.964	0.246	198	30215	APG-65	1	0	0.761	0.893
JC858	3	7	1	-50.932	162.5	0.22	348	29817	APQ-73	1	0	0.701	0.851
JC521	3	7	1	64.776	236.6	0.143	346	28564	APG-73	1	0	0.484	0.677
JC843	3	7	1	112	-82.499	0.332	303	32284	25-Apr	1	0	0.875	1
JC553	3	7	1	183.58	179.03	0.323	211	28689	APQ-159	1	0	0.915	1
JC381	3	7	1	44.212	76.991	0.333	166	26922	APG-65	1	0	0.754	0.884
JC507	3	7	1	157.36	-100.81	0.32	182	26805	APG-65	0	0	0.723	0.912
JC278	3	7	1	277.35	-231.69	0.264	154	34686	APG-65	0	0	0.726	0.912
JC999	3	7	1	163.86	-144.85	0.226	210	29196	APG-145	1	0	0.31	0.446
JC733	3	7	1	214.23	223.74	0.183	349	26476	APG-73	0	0	0.482	0.653
JC242	3	7	1	198.59	-155.3	0.301	243	38910	APG-73	0	0	0.503	0.676
JC875	3	7	1	-227.14	-170.39	0.174	343	36179	APQ-159	0	0	0.81	0.98
T535	9	7	4	91.222	-51.265	0.18	63	36076	APG-73	1	0	0.322	0.472
T642	9	7	4	-192.13	-200.12	0.34	71	29000	APG-65	1	0	0.415	0.591
T537	9	7	4	82.822	-162.31	0.33	215	39000	APG-65	1	0	0.901	1
T529	9	7	4	-192.13	-35.265	0.27	165	39000	APQ-73	1	0	0.935	1
T528	9	7	4	-194.13	-32.265	0.34	39	47348	APQ-159	0	0	0.932	1
T423	9	7	4	-138.13	-51.265	0.21	121	28925	APG-145	1	0	0.562	0.675
T521	9	7	4	-211.61	-51.265	0.18	294	38693	APQ-159	1	0	0.058	0.179
T646	9	7	4	190.1	-175.47	0.21	113	36870	APG-73	1	0	0.345	0.495
T426	9	7	4	-147.13	-105.45	0.18	8	29258	APG-65	1	0	0.113	0.281
T536	9	7	4	81.222	-51.265	0.18	14	37204	APG-65	1	0	0.13	0.243
T742	9	7	4	242.64	-24.112	0.34	332	39220	APQ-159	1	0	0.95	1
T647	9	7	4	208.43	-177.97	0.18	323	34560	APG-73	1	0	0.364	0.482
T421	9	7	4	-168.13	-51.265	0.2	99	25031	APG-145	1	0	0.527	0.699
T539	9	7	4	89.502	-14.112	0.18	54	39677	APG-65	1	0	0.091	0.268
T753	9	7	4	226.76	-150.7	0.3	350	35000	APG-73	0	0	0.946	1
T413	9	7	4	-198.13	19.265	0.34	95	39000	APG-65	1	0	0.912	1

Scenario 3

		Identity		Position						Radar		Conf. Interval	
ID	Type	curr	true	x	y	Speed	Heading	Altitude	Radar	on/off	sel	min	max
JC907	3	7	1	-134.248	75.19	0.156	187	30974	25-Apr	0	0	0.877	1
JC939	3	7	1	-58.037	-18.0136	0.264	46	33337	APG-73	1	0	0.523	0.673
JC129	3	7	1	-211.832	-7.14896	0.188	313	38843	APQ-159	0	0	0.823	0.958
JC875	3	7	1	-218.241	-63.0131	0.269	187	2245	APG-73	0	0	0.919	1
JC203	3	7	1	143.587	1.3657	0.123	318	36736	APG-73	1	0	0.931	1
JC516	3	7	1	6.89413	-176.664	0.139	296	28540	25-Apr	1	0	0.876	1
JC749	3	7	1	93.9085	141.171	0.308	306	39904	APG-145	0	0	0.902	1
JC237	3	7	1	67.1743	90.1135	0.206	165	36311	APG-65	1	0	0.763	0.916
JC602	3	7	1	-88.5434	-227.096	0.104	32	27848	APQ-159	0	0	0.918	1
JC403	3	7	1	200.443	-37.3623	0.173	64	28378	APQ-159	0	0	0.82	0.98
JC633	3	7	1	208.316	100.673	0.095	316	30138	APQ-159	1	0	0.931	1
JC613	3	7	1	89.7885	-108.73	0.127	76	39488	APQ-159	0	0	0.947	1
JC519	3	7	1	22.5684	-36.2941	0.321	62	25627	25-Apr	0	0	0.893	1
JC237	3	7	1	15.0792	203.261	0.213	281	32253	25-Apr	1	0	0.882	1
JC413	3	7	1	-214.78	159.909	0.097	262	35988	APQ-159	1	0	0.914	1
JC857	3	7	1	-22.9347	-242.813	0.085	125	30715	25-Apr	1	0	0.936	1
JC003	3	7	1	-142.396	236.511	0.316	123	28583	APQ-159	0	0	0.818	0.947
JC273	3	7	1	6.80258	-158.322	0.148	108	39477	APG-65	1	0	0.74	0.907
JC447	3	7	1	-21.8726	-118.48	0.149	280	34118	APQ-159	1	0	0.822	0.983
JC336	3	7	1	277.349	226.623	0.086	136	20183	APQ-159	1	0	0.925	1
JC309	3	7	1	-127.473	-153.531	0.237	248	29806	APG-73	1	0	0.494	0.647
JC122	3	7	1	73.0705	115.291	0.157	118	31686	APG-65	1	0	0.759	0.915
JC009	3	7	1	267.003	223.006	0.033	98	32704	APG-145	1	0	0.274	0.395
JC852	3	7	1	-129.066	-170.072	0.325	25	28384	25-Apr	0	0	0.883	1
JC278	3	7	1	-205.625	-214.217	0.245	337	26257	APG-65	0	0	0.748	0.879
JC742	3	7	1	214.469	-24.8802	0.254	255	31284	APQ-73	1	0	0.434	0.567
JC900	3	7	1	296.796	166.486	0.235	137	35114	APG-65	0	0	0.766	0.909
JC825	3	7	1	-97.3876	93.2417	0.031	36	34180	APG-65	0	0	0.915	1
JC921	3	7	1	-212.217	-191.435	0.122	278	15865	APQ-159	1	0	0.934	1
JC156	3	7	1	40.0739	29.3359	0.273	219	34853	25-Apr	1	0	0.864	1
JC274	3	7	1	165.267	47.1129	0.29	343	30570	25-Apr	0	0	0.894	1
JC507	3	7	1	217.014	-180.815	0.294	343	24456	APG-73	0	0	0.943	1
JC759	3	7	1	91.3999	-119.625	0.219	74	20942	APG-73	1	0	0.935	1
JC258	3	7	1	226.334	-234.756	0.242	170	25273	APG-65	1	0	0.728	0.889
JC929	3	7	1	299.707	-150.25	0.311	205	38140	25-Apr	0	0	0.871	1
T419	9	7	4	-192.132	28.598	0.18	340	28656	APG-73	1	0	0.361	0.49
T531	9	7	4	82.222	-168.312	0.32	222	40000	APQ-159	1	0	0.922	1
T754	9	7	4	218.205	-170.468	0.33	179	36268	25-Apr	0	0	0.922	1
T631	9	7	4	208.431	28.598	0.35	222	39000	APG-65	1	0	0.901	1
T760	9	7	4	209.65	-180.468	0.32	210	40000	APQ-159	1	0	0.924	1
T412	9	7	4	-194.132	20.265	0.34	237	37316	APG-145	0	0	0.91	1
T634	9	7	4	263.417	-51.265	0.17	332	3995	APG-145	1	0	0.608	0.76
T743	9	7	4	247.525	-26.112	0.22	293	20469	APG-145	1	0	0.611	0.72
T425	9	7	4	-144.132	-88.911	0.1	212	25423	APG-145	1	0	0.579	0.754
T538	9	7	4	87.322	-162.312	0.23	74	26187	APG-145	1	0	0.516	0.712
T645	9	7	4	171.774	-172.968	0.08	40	34189	APG-145	1	0	0.572	0.744
T637	9	7	4	208.431	16.897	0.18	174	11129	APG-65	1	0	0.125	0.284
T648	9	7	4	226.76	-180.468	0.17	10	37856	APG-73	1	0	0.345	0.477
T420	9	7	4	-98.132	30.598	0.18	96	34628	APG-65	1	0	0.13	0.272
T748	9	7	4	-192.132	-175.468	0.18	217	28713	APG-65	1	0	0.098	0.268

Scenario 4

		Identity		Position						Radar		Conf. Interval	
ID	Type	curr	true	x	y	Speed	Heading	Altitude	Radar	on/off	sel	min	max
JC907	3	7	1	67.5588	-64.982	0.215	40	2950	APQ-159	1	0	0.931	1
JC192	3	7	1	-287.07	-163.34	0.25	25	3927	APQ-73	1	0	0.701	0.879
JC821	3	7	1	-204.93	-248.55	0.181	55	22628	APQ-159	1	0	0.916	1
JC281	3	7	1	150.618	-64.508	0.215	324	38488	APG-73	0	0	0.514	0.639
JC759	3	7	1	82.7204	192.091	0.33	312	29682	APQ-73	0	0	0.703	0.881
JC237	3	7	1	-18.686	59.7629	0.148	326	33478	APG-73	0	0	0.508	0.644
JC255	3	7	1	-163.33	-60.892	0.038	194	27070	APG-73	1	0	0.93	1
JC189	3	7	1	-53.038	11.5284	0.145	145	27436	APQ-159	1	0	0.908	1
JC381	3	7	1	228.44	-128.05	0.194	266	29279	25-Apr	1	0	0.856	1
JC009	3	7	1	-123.26	45.0224	0.316	290	35417	APQ-159	1	0	0.825	0.976
JC202	3	7	1	-290.53	189.36	0.253	357	26295	APG-65	0	0	0.754	0.875
JC602	3	7	1	-78.069	188.307	0.029	333	34992	APG-65	1	0	0.938	1
JC381	3	7	1	-225.53	148.251	0.279	6	33542	25-Apr	0	0	0.86	1
JC843	3	7	1	-121.98	-101.34	0.301	335	26392	APQ-159	1	0	0.812	0.959
JC203	3	7	1	238.475	-240.75	0.1	301	21041	25-Apr	1	0	0.934	1
JC602	3	7	1	-274.91	76.7922	0.042	116	33306	APQ-159	1	0	0.938	1
JC612	3	7	1	175.228	-111.64	0.222	159	28878	25-Apr	1	0	0.862	1
JC579	3	7	1	-204.43	-22.057	0.317	162	26389	25-Apr	1	0	0.853	1
JC126	3	7	1	73.5282	170.774	0.141	46	17941	APG-65	1	0	0.708	0.883
JC891	3	7	1	-92.718	-10.308	0.042	350	17344	APQ-159	1	0	0.92	1
JC291	3	7	1	-18.43	-189.82	0.047	142	34053	APQ-159	1	0	0.927	1
JC519	3	7	1	204.27	-8.7359	0.226	48	11276	APG-145	1	0	0.273	0.415
JC019	3	7	1	168.087	246.735	0.172	140	3447	APQ-159	1	0	0.941	1
JC911	3	7	1	-133.33	-179.76	0.298	125	49685	APG-65	1	0	0.937	1
JC129	3	7	1	131.904	34.051	0.176	265	32208	APG-65	1	0	0.745	0.886
JC491	3	7	1	-104.47	-151.65	0.212	201	17715	APQ-159	0	0	0.93	1
T524	9	7	4	-178.13	-66.134	0.31	146	37000	25-Apr	1	0	0.921	1
T424	9	7	4	-148.13	-72.368	0.08	220	17980	APQ-73	1	0	0.572	0.736
T651	9	7	4	227.978	-18.112	0.22	53	30459	APG-73	1	0	0.359	0.48
T540	9	7	4	213.654	-14.112	0.17	98	4997	APG-145	1	0	0.6	0.733
T416	9	7	4	-189.13	22.265	0.23	83	26052	APG-145	1	0	0.55	0.705
T411	9	7	4	-200.13	21.265	0.08	121	38175	APG-145	1	0	0.582	0.73
T650	9	7	4	223.091	-16.112	0.19	95	26060	APG-73	1	0	0.324	0.471
T755	9	7	4	209.65	-170.47	0.36	56	39000	APQ-73	0	0	0.915	1
T752	9	7	4	220.655	-170.47	0.31	130	39000	APG-73	0	0	0.924	1
T758	9	7	4	236.76	-175.47	0.31	33	37251	APQ-159	0	0	0.935	1
T641	9	7	4	-194.13	-105.84	0.32	25	38787	APG-73	0	0	0.911	1
T534	9	7	4	72.222	30.598	0.2	358	38000	APQ-73	1	0	0.412	0.597

Scenario 5

		Identity		Position						Radar		Conf. Interval	
ID	Type	curr	true	x	y	Speed	Heading	Altitude	Radar	on/off	sel	min	max
JC003	3	7	1	195.517	-96.5377	0.274	212	32405	APQ-73	1	0	0.425	0.59
JC422	3	7	1	-52.3606	57.5961	0.104	189	31468	APG-65	0	0	0.948	1
JC281	3	7	1	220.988	-52.4995	0.066	8	27587	25-Apr	1	0	0.914	1
JC602	3	7	1	233.512	-165.403	0.059	207	29630	APG-145	0	0	0.253	0.409
JC278	3	7	1	-141.481	-49.5087	0.189	338	38238	APQ-159	1	0	0.801	0.94
JC281	3	7	1	113.831	-205.046	0.147	53	25831	25-Apr	0	0	0.895	1
JC613	3	7	1	-156.642	199.034	0.27	27	27168	APQ-159	0	0	0.793	0.96
JC907	3	7	1	-136.94	-238.861	0.153	213	35115	APQ-159	1	0	0.814	0.974
JC024	3	7	1	7.97449	-247.803	0.289	67	29977	25-Apr	1	0	0.861	1
JC032	3	7	1	-158.126	20.8365	0.173	50	28066	APQ-159	0	0	0.801	0.961
JC203	3	7	1	54.1002	22.2709	0.105	286	28329	APG-73	1	0	0.938	1
JC875	3	7	1	-287.951	-15.9993	0.223	259	36924	APQ-159	1	0	0.832	0.944
JC759	3	7	1	217.417	203.108	0.138	2	37171	APG-73	1	0	0.926	1
JC960	3	7	1	118.537	132.931	0.305	40	35016	APQ-159	1	0	0.823	0.98
JC381	3	7	1	174.844	186.201	0.41	41	29771	APQ-159	1	0	0.444	0.568
JC709	3	7	1	-245.927	141.293	0.287	265	37132	APQ-159	1	0	0.837	0.982
JC627	3	7	1	220.054	-98.5366	0.278	179	32929	25-Apr	1	0	0.89	1
JC992	3	7	1	40.9162	-92.5245	0.116	120	35271	APG-73	0	0	0.946	1
JC615	3	7	1	10.0253	235.794	0.284	261	27636	APG-65	1	0	0.764	0.886
JC553	3	7	1	-10.9592	-93.4706	0.286	204	34112	APG-65	0	0	0.718	0.918
JC242	3	7	1	-54.21	-95.1338	0.248	3	29736	APG-145	1	0	0.328	0.459
JC049	3	7	1	143.806	77.4789	0.283	47	26501	APG-145	1	0	0.262	0.391
JC139	3	7	1	-289.123	184.736	0.259	223	29695	APQ-73	0	0	0.407	0.589
JC202	3	7	1	-239.775	-31.0755	0.255	272	25869	APG-65	0	0	0.744	0.912
JC947	3	7	1	215.934	-74.5949	0.055	130	15539	APG-145	1	0	0.404	0.561
JC179	3	7	1	-43.2783	64.8595	0.274	313	27734	APG-73	0	0	0.944	1
JC885	3	7	1	-192.77	150.006	0.093	265	35075	APG-73	1	0	0.93	1
T418	9	7	4	-194.132	26.598	0.31	39	45487	APQ-159	0	0	0.919	1
T745	9	7	4	257.299	-16.112	0.36	226	35654	APG-65	0	0	0.902	1
T750	9	7	4	-144.118	-180.468	0.19	80	31659	APG-145	0	0	0.523	0.693
T541	9	7	4	171.774	-168.312	0.18	325	26412	APQ-159	1	0	0.017	0.183
T749	9	7	4	-168.125	-177.968	0.18	51	35476	APQ-159	1	0	0.05	0.204
T636	9	7	4	190.102	-16.112	0.09	67	29539	APG-145	1	0	0.436	0.585
T640	9	7	4	-179.132	-200.122	0.37	5	35243	APG-65	0	0	0.915	1
T638	9	7	4	226.76	-16.112	0.22	347	32553	APG-145	0	0	0.549	0.67
T527	9	7	4	-179.132	-31.265	0.19	207	38000	APG-73	1	0	0.321	0.484
T630	9	7	4	190.102	-168.312	0.32	5	40000	APQ-159	0	0	0.923	1
T751	9	7	4	220.655	-170.468	0.34	31	40619	25-Apr	0	0	0.9	1
T530	9	7	4	-98.132	-31.265	0.21	350	5134	APG-145	1	0	0.576	0.762
T649	9	7	4	218.205	-14.112	0.36	49	39000	APG-65	0	0	0.921	1
T741	9	7	4	237.752	-22.112	0.07	31	36383	APQ-73	1	0	0.597	0.722
T635	9	7	4	171.774	-105.84	0.07	18	22209	APQ-73	1	0	0.572	0.736

Scenario 6

		Identity		Position						Radar		Conf. Interval	
ID	Type	curr	true	x	y	Speed	Heading	Altitude	Radar	on/off	sel	min	max
JC615	3	7	1	-190.793	-141.156	0.17	145	7122	25-Apr	1	0	0.91	1
JC003	3	7	1	-203.775	48.04376	0.098	29	37819	APG-65	0	0	0.929	1
JC189	3	7	1	-147.89	-26.7113	0.057	212	33478	APQ-159	1	0	0.922	1
JC299	3	7	1	-133.534	30.32777	0.217	195	36065	APG-73	1	0	0.937	1
JC291	3	7	1	-242.686	-87.3211	0.256	347	37058	25-Apr	0	0	0.872	1
JC242	3	7	1	-57.3962	125.1946	0.553	143	35549	APG-73	1	0	0.929	1
JC602	3	7	1	131.2082	-132.946	0.495	243	38739	APG-73	1	0	0.931	1
JC443	3	7	1	-204.599	247.2381	0.177	137	26775	APG-65	0	0	0.703	0.862
JC843	3	7	1	-193.503	133.343	0.208	265	33262	APG-65	0	0	0.756	0.893
JC875	3	7	1	-23.6671	32.09784	0.33	126	26118	APG-145	0	0	0.301	0.465
JC939	3	7	1	258.9831	-208.754	0.23	50	33321	APG-73	0	0	0.529	0.646
JC129	3	7	1	216.7394	62.82998	0.22	91	29294	APG-73	1	0	0.929	1
JC097	3	7	1	3.140355	-0.45015	0.217	346	31360	APG-145	1	0	0.292	0.478
JC247	3	7	1	289.9289	-52.7741	0.2	110	39906	APG-73	0	0	0.481	0.65
JC311	3	7	1	152.5224	103.6637	0.209	202	48528	25-Apr	0	0	0.927	1
JC003	3	7	1	-290.46	-32.8455	0.183	201	38968	APQ-159	1	0	0.798	0.962
JC742	3	7	1	-177.535	110.7746	0.148	77	27120	APQ-159	1	0	0.804	0.948
JC615	3	7	1	-5.6856	-160.764	0.239	319	47511	APQ-73	1	0	0.43	0.584
JC633	3	7	1	247.8133	-120.891	0.195	341	17570	APQ-73	1	0	0.747	0.893
JC759	3	7	1	261.9312	198.7747	0.034	101	32565	APQ-159	1	0	0.945	1
JC553	3	7	1	-143.074	83.67412	0.437	355	32218	APG-145	1	0	0.929	1
JC839	3	7	1	43.55296	-117.87	0.167	56	31400	APG-73	1	0	0.925	1
JC613	3	7	1	-165.816	-22.6066	0.144	87	25843	APG-73	1	0	0.502	0.647
JC022	3	7	1	-150.233	-202.406	0.207	61	30058	APG-65	1	0	0.766	0.901
JC639	3	7	1	-165.835	-83.6283	0.166	35	38986	APG-65	1	0	0.724	0.897
JC834	3	7	1	-145.747	-2.266	0.314	74	36801	APG-65	1	0	0.722	0.872
JC111	3	7	1	120.2765	-11.6047	0.225	188	35397	25-Apr	1	0	0.936	1
JC409	3	7	1	-97.2594	-219.802	0.174	67	39109	APG-145	1	0	0.323	0.48
JC922	3	7	1	100.5188	-220.87	0.293	108	17809	APG-65	1	0	0.725	0.854
JC629	3	7	1	31.87048	-179.044	0.189	282	38799	APG-73	1	0	0.484	0.637
JC202	3	7	1	260.9058	153.8667	0.172	77	30773	APQ-159	1	0	0.83	0.966
JC947	3	7	1	215.9337	-74.5949	0.071	130	26240	APQ-159	1	0	0.941	1
T756	9	7	4	-168.125	-200.122	0.33	121	36655	APG-65	0	0	0.945	1
T522	9	7	4	-219.494	-31.265	0.31	59	38918	APQ-159	0	0	0.911	1
T428	9	7	4	-198.132	-138.537	0.23	278	29555	APG-73	0	0	0.337	0.482
T533	9	7	4	81.222	28.598	0.18	98	33590	25-Apr	1	0	0	0.135
T757	9	7	4	220.655	-170.468	0.34	348	43977	25-Apr	0	0	0.907	1
T643	9	7	4	-168.125	-105.84	0.17	269	31925	APG-73	0	0	0.335	0.516
T523	9	7	4	-227.38	-31.265	0.06	51	37019	APQ-73	1	0	0.615	0.743
T632	9	7	4	226.76	30.598	0.31	169	45667	APG-65	0	0	0.904	1
T427	9	7	4	-139.132	-121.995	0.19	248	38000	APQ-73	1	0	0.42	0.578
T525	9	7	4	-188.132	-79.989	0.38	116	34000	APQ-159	1	0	0.402	0.574
T644	9	7	4	220.655	-170.468	0.08	115	34573	APG-145	1	0	0.604	0.758
T520	9	7	4	-203.722	-51.265	0.18	266	31919	APG-65	1	0	0.107	0.274
T747	9	7	4	-194.132	-200.122	0.2	112	38390	APQ-73	0	0	0.417	0.596

APPENDIX C EXPERIMENTAL MATERIALS

Task Instructions (Non-Intentional Condition)

In this investigation, you will perform some of the duties of an Advanced Warning Aircraft Communication (AWAC) system operator. It is your responsibility to monitor information provided to you on the computer terminal regarding the location, capabilities, and characteristics of friendly and threatening aircraft, and to use that information to protect friendly assets by identifying any forces that are unknown to you. Other personnel will use that information to take action against hostile forces that you have identified.

Currently, your control center is operating in the area of the Italy and the Mediterranean Seam including the Tyrrhenian, Ionian, and Adriatic Seas. You are responsible for identifying all unknown aircraft in this area.

To accomplish this, you have the following capabilities:

1. Request reconnaissance and electronic sensor information.
2. Labeling unknown forces as threatening or friendly.

To perform your task, you are given a variety of electronic information sources, including an automated decision aid, which provides you with a probabilistic estimate of an aircraft's identity. This system is based on a variety of electronic and intelligence sources of information, and is combined using advanced mathematical techniques. Past experience has shown that this aid helps commanders make identification decisions, but may be subject to occasional hardware or software problems which may cause the aid to produce unreliable estimates.

Due to limited resources, the best decisions will be those that minimize the time consuming requests for information. However, it is imperative to identify aircraft correctly, so that friendly forces are not inadvertently attacked, and so that friendly assets can be protected. Therefore, your score for each session will be based on both your correct identifications, and the amount of information you request to make identifications. Specifically, you will be penalized 10 points for every object you misidentified. Total of 600 points, the correct identification contributes 500 points. You will also be penalized 1 point for every time you open up the decision aid windows. You will be able to see your score at the end of each session.

Task Instructions (Sabotage Condition)

In this investigation, you will perform some of the duties of an Advanced Warning Aircraft Communication (AWAC) system operator. It is your responsibility to monitor information provided to you on the computer terminal regarding the location, capabilities, and characteristics of friendly and threatening aircraft, and to use that information to protect friendly assets by identifying any forces that are unknown to you. Other personnel will use that information to take action against hostile forces that you have identified.

Currently, your control center is operating in the area of the Italy and the Mediterranean Seam including the Tyrrhenian, Ionian, and Adriatic Seas. You are responsible for identifying all unknown aircraft in this area.

To accomplish this, you have the following capabilities:

1. Request reconnaissance and electronic sensor information.
2. Labeling unknown forces as threatening or friendly.

To perform your task, you are given a variety of electronic information sources, including an automated decision aid, which provides you with a probabilistic estimate of an aircraft's identity. This system is based on a variety of electronic and intelligence sources of information, and is combined using advanced mathematical techniques. Past experience has shown that this aid helps commanders make identification decisions, but may be subject to intentional interference with the computer system by enemy forces which may cause the aid to produce unreliable estimates.

Due to limited resources, the best decisions will be those that minimize the time consuming requests for information. However, it is imperative to identify aircraft correctly, so that friendly forces are not inadvertently attacked, and so that friendly assets can be protected. Therefore, your score for each session will be based on both your correct identifications, and the amount of information you request to make identifications. Specifically, you will be penalized 10 points for every object you misidentified. Total of 600 points, the correct identification contributes 500 points. You will also be penalized 1 point for every time you open up the decision aid windows. You will be able to see your score at the end of each session.

Task Instructions (Control Condition)

In this investigation, you will perform some of the duties of an Advanced Warning Aircraft Communication (AWAC) system operator. It is your responsibility to monitor information provided to you on the computer terminal regarding the location, capabilities, and characteristics of friendly and threatening aircraft, and to use that information to protect friendly assets by identifying any forces that are unknown to you. Other personnel will use that information to take action against hostile forces that you have identified.

Currently, your control center is operating in the area of the Italy and the Mediterranean Seam including the Tyrrhenian, Ionian, and Adriatic Seas. You are responsible for identifying all unknown aircraft in this area.

To accomplish this, you have the following capabilities:

1. Request reconnaissance and electronic sensor information.
2. Labeling unknown forces as threatening or friendly.

To perform your task, you are given a variety of electronic information sources, including an automated decision aid, which provides you with a probabilistic estimate of an aircraft's identity. This system is based on a variety of electronic and intelligence sources of information, and is combined using advanced mathematical techniques. Past experience has shown that this aid helps commanders make identification decisions.

Due to limited resources, the best decisions will be those that minimize the time consuming requests for information. However, it is imperative to identify aircraft correctly, so that friendly forces are not inadvertently attacked, and so that friendly assets can be protected. Therefore, your score for each session will be based on both your correct identifications, and the amount of information you request to make identifications. Specifically, you will be penalized 10 points for every object you misidentified. Total of 600 points, the correct identification contributes 500 points. You will also be penalized 1 point for every time you open up the decision aid windows. You will be able to see your score at the end of each session.

Rules of Engagement

Assess Situation and Update Air Data

1. *Perform target identification of all unknown air tracks* before tracks leave the area covered by your radar scope

Sources of Information are:

a. Radar Signature: The Electronic Warfare Supervisor provides sensor (radar) information on any requested track.

b. Altitude: Radar System provides altitude information on any track by request.

c. Speed: Radar System provides speed information on any track by request.

d. Track Confidence Interval: (shown in the data fusion window) This interval indicates how much the system is confident on the object interested being friendly. Thus, the higher the number and the narrower the interval, the more likely to be a friendly object.

(You will be provided with the heading information, also)

2. *Modify unknown air track parameters on console.*

NOTE: Other officers are responsible for warning, illuminating, and/or firing upon tracks identified as hostile as they approach to our region. Mis-identification of tracks therefore has grievous consequences. Unknown tracks cannot be warned or engaged and therefore pose a threat to our allies if they approach within certain range and are not identified.

Past experience has indicated that aircraft with the following characteristics have different chances of *being friendly*.

1. Sensor Summary

Sensor	**Aircraft**	**Base probability**
APG-145	Hawkeye	0.40
APG-73	F/A-18	0.70
APQ-73	ATA*	0.40
APG-65	Tornado	0.70
APQ-159		0.80
APR-25		0.75

*ATA: Genotype of Advanced Tactical Aircraft

2. Altitude Profile Summary

Profile (feet)	**Base Probability**
100 – 30,000	0.80
25,000 – 40,000	0.75
35,000 – 50,000	0.20

** Your aircraft cannot detect an object flying under 100 feet high.

3. Speed Profile Summary

Profile (mile per hour)	**Base Probability**
0 – 600	0.80
500 – 1200	0.75
1100 and over	0.20

You will perform this task using a personal computer equipped with a keyboard, monitor, and mouse. Your actions (e.g., mouse clicks and keystrokes) will be automatically recorded by the computer during the experiment.

Today, you will first practice using the computer by performing a short (10 minute) sample task, during which the experimenter will show you how to use the mouse and keyboard to run the program. If you have any questions, at this time or throughout the experiment, please ask the experimenter. If you want to practice the task more than once, please ask the experimenter. After you feel comfortable using the program, you will perform two experimental scenarios lasting 20 minutes each. For the rest of the days of the experiment, you will perform two experimental scenarios lasting 20 minutes each.

Participant Data Form

Subject # ____________

Age: ________

Sex: M F

Full time semesters of undergraduate school:_________

Full time semesters of graduate school:_________

Undergraduate major, minor:

Graduate major, minor:

Please list any probability and statistics or decision theory courses you have taken, or other similar experience:

Please briefly describe any work, hobby, or volunteer experience you have had (e.g., military experience) which may have influenced your performance in the experiment:

Debriefing

The purpose of this study was to examine how people make choices in dynamic tasks, and how they decide whether or not to use automated or computerized decision aids. Specifically, we are trying to understand under what conditions people notice when information provided by computerized decision aids becomes unreliable, or contains error. In the experiment you performed, we wanted to test whether telling people that a decision aid could fail due to hardware or software failures, or because an enemy intentionally corrupts the information, has an effect on how people use and rely on the decision aid. Additionally, we want to understand how the type of failure – for example, the gradual addition of noise to the information so that the aid provides less exact estimates, or intermittent breakdowns in which the aid failed to provide the information it should have, affected the way in which the aid was used.

More information regarding the use of automated aids can be found in:

Lee, J. D., Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. Ergonomics, 35(10), 1243-1270.

Lee, J. D. and Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. International Journal of Human-Computer Studies, 40, 153-184.

More information regarding decision making in complex situations can be found by reading:

Zsambok, C. E. and Klein, G. (1997). Naturalistic Decision Making. Mahwah, NJ: Lawrence Erlbaum.

If you have further questions about the experiment, or wish to learn more about the study of human decision making, please contact

Dr. Ann Bisantz
414 Bell Hall
645-2357 x 2474
bisantz@eng.buffalo.edu

APPENDIX D EXPERIMENTAL TESTBED DESCRIPTION OF FILES

Class name	Description
DataFusion	This class generates the data fusion window to simply show the confidence interval for the selected track.
Entity	Entity class defines all the contacts of tracks, such as speed, heading, radar turning on/off.
Event	This class controls the time of track displaying on the main window, and turning the radar on and off.
FileParser	FileParser class is used only to load tracks and events.
MapDemoDoc	This class is to display the map and to call a scenario file.
MapDemoView	Main simulation,is controlled by this class. Not only does it control what to display on the screen, but also this class handles the command input from the users. Simply, this class possesses the thread.
Performance	Performance feedback was given at the end of each simulation.
StatusSheet	This shows all contacts (speed, heading, altitude, and radar signature) of the selected track.
TrustQuestion	This class shows the electronic version of questionnaire developed earlier in Chapter 2 (see also Figure 4.4).
Warning	This class is used only in the third trial to inform users about errors.

GLOSSARY

AADM	Aided-adversarial Decision-making
IW	Information Warfare
JDL/DFG	Joint Directors of Laboratories Data Fusion Group
RPD	Recognition Primed Decision
TIW	Track Information Window
DFW	Data Fusion Window